



**UNIVERSIDADE FEDERAL DE SERGIPE**  
**CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE – CCBS**  
**DEPARTAMENTO DE BIOLOGIA – DBI**

**EVERTON SILVA MOTA**

**BIOINFORMÁTICA NO ENSINO DE GENÉTICA PARA O CURSO DE  
GRADUAÇÃO DE CIÊNCIAS BIOLÓGICAS SOB METODOLOGIA ATIVA**

**SÃO CRISTÓVÃO**

**2018**

**EVERTON SILVA MOTA**

**BIOINFORMÁTICA NO ENSINO DE GENÉTICA PARA O CURSO DE  
GRADUAÇÃO DE CIÊNCIAS BIOLÓGICAS SOB METODOLOGIA ATIVA**

Monografia apresentada ao Curso de Licenciatura em Ciências Biológicas do Departamento de Biologia da Universidade Federal de Sergipe como requisito parcial para obtenção do título de Licenciado em Ciências Biológicas.

Orientador: Prof. Dr. Marcus Vinicius de Aragão Batista

**SÃO CRISTÓVÃO**

**2018**

## **AGRADECIMENTOS**

A parte mais difícil da monografia, mas lá vai.... Gostaria de agradecer primeiramente a Deus, por me dar forças e me manter firme até aqui, em seguida aos meus pais que sempre me forneceram tudo que eu precisei para que eu conseguisse continuar estudando e realizando os meus sonhos, Evando e Joelma, sem vocês eu não seria ninguém.

Ao meu irmão nas vezes que mesmo brincando dizia: “vá estudar rapaz” porque as vezes a preguiça chega e quer fazer morada, obrigado Erivan pelo apoio.

Gostaria de agradecer ao meu Orientador Dr. Marcus Vinicius de Aragão Batista, pela paciência (e tem que ter muita comigo), orientações, aconselhamentos e até as broncas que são necessárias, obrigado Marcus, você contribuiu muito na minha formação profissional e pessoal.

Ao Dr. Edilson Divino de Araújo, pois se não fosse por você, eu nem estaria no GMBio, muito obrigado por abrir as portas da ciência para mim, pelos conselhos, paciência e por toda confiança depositada, você faz parte da minha história.

À professora Dr<sup>a</sup>. Luciane Storti, com seus aconselhamentos sobre os projetos (e as vezes sobre a vida) muito obrigado, sei que nos conhecemos a pouco tempo, mas você já contribuiu bastante.

A todos os meus professores, da graduação e do ensino fundamental e médio, vocês formaram esse cidadão, e algumas frases que ouvi no ensino fundamental me inspiram e motivam até hoje, obrigado!

Ao pessoal do laboratório GMBio, (momento do medo de esquecer o nome de alguém) Conceição, Débora, Edilaine, Gerlane, Jordana, Lidiane, Fernanda, Luana, “Lucas<sup>2</sup>”, Melina, Rebeca, Myrela e Genilda; vocês me ensinaram muita coisa, e como sempre dizemos, a gente ganha pouco, mas se diverte! É muito bom trabalhar e ser amigo de vocês.

À minha turma, (de loucos) vocês sabem que sem vocês o caminho teria sido muito mais árduo, vocês me ajudaram muito, em vários aspectos, muito obrigado pela parceria de sempre, em especial à Tanísia e Geovane que me ajudaram na validação do questionário, muito obrigado!

Resumindo, a todos que me ajudaram direta ou indiretamente, muitíssimo obrigado!

## **Resumo**

Com o avanço da Biologia Molecular surgiram vários conjuntos de dados e, junto com eles, houve a necessidade de sistematizar e analisar esses dados para a geração de informações. Para desempenhar tal função surge uma nova área que engloba as Ciências Biológicas, Matemática e Ciências da Computação: A Bioinformática. Poucas universidades brasileiras oferecem cursos que lidam com aspectos relacionados à bioinformática em nível de graduação, assim, muitos biólogos deixam universidades sem ter nenhum conhecimento sobre Bioinformática. Por ser uma área intimamente ligada à Biologia Molecular, a Bioinformática se mostra uma importante ferramenta para auxiliar no ensino de algumas disciplinas, como a Genética. Neste contexto, o presente trabalho tem como objetivo introduzir a disciplina de Bioinformática para graduandos do curso de Ciências Biológicas matriculados na disciplina de Genética Molecular, através de uma metodologia ativa de ensino, visando avaliar a relevância e importância dessa disciplina para alunos de graduação, bem como a eficácia desta metodologia de ensino. Verificou-se que os próprios alunos sentem a necessidade de uma abordagem mais aprofundada da Bioinformática no currículo acadêmico, além disso, observa-se que metodologias ativas auxiliam no processo de ensino-aprendizagem desses alunos, facilitando uma aprendizagem significativa.

**Palavras-chave:** Ensino de Biologia; Genética Molecular; Metodologia ativa; Inovação Curricular.

## **Abstract**

With the advancement of Molecular Biology several datasets have emerged and together with them, there was a need to systematize and analyze this data for the generation of information. To perform such a function arises a new area that encompasses the Biological Sciences, Mathematics and Computer Science: The Bioinformatics. Few Brazilian universities offer courses that deal with aspects related to bioinformatics at the undergraduate level, thus, many biologists leave universities without having any knowledge about Bioinformatics. Because it is an area closely linked to Molecular Biology, Bioinformatics proves to be an important tool to aid in the teaching of some disciplines such as Genetics. In this context, the present work aims to introduce the discipline of Bioinformatics for undergraduates of the Biological Sciences course enrolled in the discipline of Molecular Genetics, through an active teaching methodology, aiming to evaluate the relevance and importance of this discipline for undergraduates, as well as the effectiveness of this teaching methodology. It was verified that the students themselves feel the need for a more in-depth approach to Bioinformatics in the academic curriculum, besides, it is observed that active methodologies assist in the teaching-learning process of these students, facilitating a meaningful learning.

**Key words:** Teaching of Biology; Molecular genetics; Active methodology; Curricular Innovation.

## **Lista de figuras**

Figura 01 – Gráfico demonstrando cenário atual na oferta da disciplina de bioinformática nos cursos de Ciências Biológicas em Universidades Públicas brasileiras.....	13
---	----

## **Lista de tabelas**

Tabela 01 - Temas das aulas e horas /aula dedicada a cada tema.....	22
Tabela 02 - Distribuição dos temas para os grupos formados.....	23

## **Abreviações**

DNA – Ácido Desoxirribonucleico;

PGH – Projeto Genoma Humano;

ZDP – Zona de desenvolvimento Proximal;

ZDR – Zona de Desenvolvimento Real;

TCLE – Termo de consentimento Livre e Esclarecido.



## Sumário

<b>1- INTRODUÇÃO.....</b>	<b>10</b>
<b>2- REFERENCIAL TEÓRICO.....</b>	<b>11</b>
2.1- INÍCIO DA BIOINFORMÁTICA .....	11
2.2 - CENÁRIO ATUAL DA BIOINFORMÁTICA .....	12
<b>2.3 – BIOINFORMÁTICA COMO RECURSO DIDÁTICO .....</b>	<b>14</b>
2.4 – PRÁXIS NA EDUCAÇÃO .....	15
<b>3 - OBJETIVOS .....</b>	<b>17</b>
3.1 – GERAL:.....	17
3.2 – ESPECÍFICOS: .....	17
<b>4 - METODOLOGIA .....</b>	<b>18</b>
4.1 - ÁREA DE ESTUDOS .....	18
4.2 - PROCEDIMENTO TÉCNICO .....	18
4.3 - COLETA DE DADOS .....	21
<b>5 - RESULTADOS E DISCUSSÃO .....</b>	<b>22</b>
<b>6 - CONCLUSÕES .....</b>	<b>26</b>
<b>REFERÊNCIAS .....</b>	<b>27</b>

## 1- INTRODUÇÃO

Desde a descoberta do DNA, diversas pesquisas foram propostas a fim de trazer mais informações sobre esta molécula, estas pesquisas trouxeram dados importantes que explicitam como o DNA atua nos organismos, e como é possível manipulá-lo em laboratórios. A partir dessas informações novos equipamentos surgiram, permitindo estudos mais aprofundados.

Com este avanço tecnológico, houve também o avanço da Biologia Molecular, impulsionando assim a geração de dados e informações sobre a mesma. Com o número destes dados crescendo, surge então a necessidade de organizá-los e analisá-los, gerando de fato informação científica. Para tal problemática, surge uma disciplina que procura resolver estas questões: a Bioinformática.

Nos cursos de graduação, em Ciências Biológicas por exemplo, poucas são as universidades que promovem o contato dos alunos de graduação às ferramentas de Bioinformática, tão pouco, utilizam destas ferramentas para facilitar o processo de ensino-aprendizagem nas disciplinas de graduação que mais podem ser favorecidas por essas, tais como: Genética, Biologia molecular, Evolução, entre outras.

No Brasil, é perceptível a necessidade da geração de recursos humanos capazes de atuar nesta área, uma vez que o aumento de dados moleculares gerados pelas pesquisas científicas, estabelece uma demanda significativa. Entretanto, muitos alunos da graduação saem das universidades sem sequer saber da existência da Bioinformática. Além da importância do aprendizado da Bioinformática para os alunos de Ciências Biológicas, as ferramentas computacionais da Bioinformática podem ser usadas para promover a capacidade investigativa dos alunos de graduação ao se propor problemas biológicos que possam ser solucionados com o auxílio delas, auxiliando no processo de ensino-aprendizagem de diversas disciplinas do curso, principalmente a Genética.

## 2- REFERENCIAL TEÓRICO

### 2.1- Início da Bioinformática

Desde a descoberta do DNA como molécula responsável pelo armazenamento da informação genética e a explicação da sua estrutura proposta por Watson e Crick em 1953, esta molécula se tornou o principal objeto de estudo dentro da biologia molecular. A partir dessas informações, novas tecnologias surgiram, o que permitiu um estudo mais aprofundado dos mecanismos utilizados por esta molécula para a manutenção e transmissão destas informações.(Farah, 2007).

Um importante evento que desencadeou uma série de estudos acerca do DNA foi o Projeto Genoma Humano (PGH), teve início oficialmente em 1990 tendo como alguns de seus objetivos: determinar toda a sequência de bases do genoma; identificar todos os seus genes; armazenar as informações em bancos de dados acessíveis ao público e aperfeiçoar a análise dos dados (Farah, 2007).

Inicialmente, o sequenciamento do genoma humano seria feito usando o mesmo método utilizado em 1995 para sequenciar o genoma de *Haemophilus influenzae* com 1,8 milhões de pares de bases. Neste método, o genoma era partido em fragmentos menores e remontado *a posteriori* com o auxílio de ferramentas computacionais que reconheceriam os pontos de sobreposição das sequências (Farah, 2007).

A partir do PGH, outros projetos foram desenvolvidos sequenciando genomas de plantas, mamíferos, invertebrados, fungos, bactérias, entre outros (Venter, 2010). Atualmente, diversas espécies já possuem seus genomas sequenciados, e este número continua a crescer (GenBank, 2017; GOLD, 2017).

Com o avanço das pesquisas, novos equipamentos foram projetados, visando otimizar os trabalhos acerca da Biologia Molecular, desta forma, houve um aumento na produção de dados moleculares criando uma demanda de profissionais capazes de obter, organizar e analisar estes dados., desta maneira, a Bioinformática tem se tornado uma importante ferramenta para a obtenção, organização e análise destes, atuando na solução de problemas biológicos (Nunes *et al.*, 2015).

Através da evolução dos projetos genomas e das novas tecnologias da informação durante a última década do século XX, foi condicionado o surgimento de uma disciplina

que gerou vínculos entre a informática, estatística e as ciências biológicas: a Bioinformática. Esta disciplina se encontra na interseção das ciências da vida com as ciências da informação (Perezleo *et al.*, 2003).

A Bioinformática corresponde a um campo científico interdisciplinar que se propõe a investigar e desenvolver sistemas que facilitem a compreensão do fluxo de informação desde os genes até as estruturas moleculares, como também suas funções bioquímicas, suas condutas fisiológicas e finalmente sua influência nas enfermidades e saúde (Perezleo *et al.*, 2003).

## **2.2 - Cenário atual da Bioinformática**

Existe uma frequente associação entre os termos “Bioinformática” e “Biologia Computacional”, porém, estes possuem conceitos distintos. De acordo com GEWIN (2011p.143), os bioinformatas desenvolvem novas maneiras de adquirir, organizar e analisar dados biológicos, uma vez que os biólogos computacionais desenvolvem modelos matemáticos ou técnicas de simulações para determinar o significado biológico dos dados.

Para Alves (2013) a Bioinformática atua no armazenamento, processamento, análise, previsão e modelagem de dados biológicos com a ajuda das ciências e tecnologias da computação.

A Bioinformática como ramo da ciência tem se mostrado bastante promissora, revelando-se uma nova ferramenta na agregação do conhecimento para a biologia molecular atrelado a informática. É notável que a bioinformática percorre diversos campos dentro da biologia, como no planejamento na conservação da biodiversidade, na biologia sintética, ou aspectos mais teóricos como os genes e a interação proteica, trazendo resultados que perpassam a biologia e atingem outras áreas como a medicina, farmácia, biotecnologia, entre outros (Ouzounis, 2012).

No ramo da saúde, a Bioinformática é uma importante ferramenta para a compreensão de como as informações genéticas são refletidas em características fisiológicas, como a susceptibilidade ao câncer, ou ainda o mecanismo patológico de uma mutação (Alves, 2013).

Na Farmácia, a Bioinformática é essencial na seleção de moléculas com o perfil necessário para a produção de um determinado fármaco, agregando informações como estrutura, propriedades das moléculas, entre outros (Alves, 2013).

A Bioinformática traz a necessidade de um novo profissional no mercado, este profissional deve obter conhecimento sobre problemas biológicos, sendo capaz de analisa-los a partir dos resultados observados, deve ainda obter conhecimentos sobre estatística e ciências da computação para executar análises mais robustas. Contudo este profissional ainda é escasso no mercado (Alves, 2013).

É possível observar que a inserção da Bioinformática no Brasil é raríssima, e ainda assim dá-se geralmente nos cursos de pós-graduação, havendo apenas algumas poucas universidades que introduzem os conhecimentos da Bioinformática ainda na graduação. (Ribeiro Junior, Oliveira e Ceccatto, 2012)

Em outros países como Alemanha, Estados Unidos e Reino Unido, o cenário é divergente, uma vez que cursos de Bioinformática são ofertados em diferentes níveis de graduação (Zatz, 2002; Koch; Fuellen, 2008; Counsell, 2002). Em outros como Malásia, México, Cuba e Costa Rica, há um forte investimento no ensino da Bioinformática, que ocorre, de uma maneira geral, devido a demanda que surgiu com o rápido progresso da biologia molecular e biotecnologia nestes países (Zeti *et al.*, 2009).

Portanto, sob uma análise da situação atual da Bioinformática, é perceptível a necessidade de uma melhor apresentação desta nova disciplina das ciências biológicas aos graduandos. É válido ainda afirmar que o uso de ferramentas de Bioinformática expressa grandes possibilidades de contribuir significativamente com a aprendizagem dos alunos de graduação e consequentemente na formação dos futuros profissionais.

É necessário mencionar que a bioinformática é uma área em ascensão e muito importante para os biólogos, porém, há uma defasagem no ensino desta nos cursos de graduação, logo este trabalho almeja avaliar como a introdução desta disciplina interfere na qualidade da formação dos alunos de Biologia.

## 2.3 – Bioinformática como recurso didático

As metodologias ativas de ensino são aquelas que possibilitam ao estudante construir seu conhecimento, ao se tornar o centro das discussões relacionadas ao tema. (Melo e Sant’Ana, 2013).

Sob esta perspectiva a Bioinformática traz características vantajosas quanto ao seu uso como recurso didático. Além de fornecer uma série de ferramentas que podem ser utilizadas de maneira ativa pelo aluno, ela amplia o pensamento crítico e reflexivo dos mesmos.

Com o objetivo de auxiliar na aprendizagem dos alunos, diversos autores utilizaram a Bioinformática como recurso didático.

Em Nunes *et al* (2015), os alunos utilizaram linguagem de programação para compreender processos ligados à Biologia Molecular tais como: Transcrição, Tradução, Complementariedade da fita de DNA, entre outros. Esse trabalho trouxe uma abordagem diferenciada para o ensino de Biologia Molecular.

Para Badotti *et al* (2013), a Bioinformática se mostrou promissora para o ensino de proteômica, de modo que mesmo com o auxílio de tutoriais, os alunos são mais ativos no processo e de fato pensam de maneira analítica.

De acordo com Mertz e Streu (2015), a metodologia ativa e a Bioinformática podem ser utilizadas para melhorar as habilidades de escrita e raciocínio dos alunos. Em seu trabalho, os autores propõem aos estudantes a escrita de um projeto com base em cursos de bioquímica fornecidos previamente, porém, nestes cursos os conteúdos são mediados por ferramentas de Bioinformática sob metodologia ativa, facilitando a compreensão dos processos abordados. Assim, o aluno consegue pensar em seguidas maneiras de usar as ferramentas estudadas para a criação de outros projetos de pesquisa.

Outros trabalhos têm utilizado a Bioinformática para auxiliar no processo de ensino de diversas disciplinas, tais como: Bioquímica (Inlow, Miller e Pittman, 2007; Junior, 2010; Badotti *et al.*, 2014; Mertz e Streu, 2015), Matemática (Chapman, Christmann e Thatcher, 2010; Wightman e Hark, 2012), Programação, Química, Física (Furge *et al.*, 2009), Fisiologia (Zhang, 2009; Nunes *et al.*, 2015), entre outros.

## 2.4 – Práxis na Educação

Sabe-se que a técnica está fortemente associada à teoria e o contrário também é válido (Ribeiro e Jesus, 2015). Portanto, cabe ao docente buscar associar sempre as teorias com as práticas visando auxiliar o processo de aprendizagem, uma vez que a prática ajuda até mesmo na contextualização e propõe significado aos conteúdos abordados.

Para Silva (2012), o docente deve fornecer algumas características ao discente, estas seriam:

- Aquisição de consciência criativa, participativa e questionadora;
- Apresentar referências teóricas para análise e interpretação da realidade;
- Ação educativa capaz de veicular teoria e prática.

A escolha de metodologias que facilitem este processo é de grande valia para a reflexão do professor, acerca das metodologias de ensino que este emprega nas suas aulas, estas reflexões podem auxiliar o docente em atingir tais objetivos.

Krasilchik (2008) relata que as aulas expositivas foram e continuam sendo a metodologia mais utilizada nas aulas de Biologia, tendo fundamental importância para a consolidação de alguns conceitos básicos dentro da ciência, porém, por diversas vezes é notório que não há uma aprendizagem significativa por parte dos alunos. Na busca por solucionar esta situação, alguns pesquisadores como Krasilchik (2008) e Carvalho (2013) têm proposto metodologias ativas de ensino. Nestas metodologias os alunos deixam de ser receptores de um conteúdo expositivo e passam a participar do processo de ensino-aprendizagem.

Dentre as metodologias ativas existentes, a aprendizagem por investigação tem se mostrado estimulante para os alunos e tem contribuído para a aquisição de algumas habilidades consideradas fundamentais na formação do cidadão (Moreira, Souza e Almassy, 2014).

De acordo com Baptista (2010), existem conceitos divergentes para ensino por investigação, em que alguns relacionam o termo investigação com processos científicos, outros relacionam à resolução de problemas.

Estas atividades permitem ao aluno:

[...] refletir, buscar explicações e participar com mais ou menos intensidade (dependendo da atividade didática proposta e de seus objetivos) das etapas de um processo que leve à resolução do problema proposto, enquanto o professor muda sua postura, deixando de agir como transmissor do conhecimento, passando a agir como um guia (AZEVEDO, 2004, p. 21).

A Bioinformática traz a oportunidade de uma abordagem pedagógica diferenciada através da resolução de problemas (Wightman e Hark, 2012). Utilizar de suas ferramentas para estimular o aluno na busca por soluções é um método interessante para desenvolver habilidades esperadas para os futuros profissionais.

Portanto, sob uma análise da situação atual da Bioinformática, questiona-se a necessidade da introdução desta nova disciplina das ciências biológicas aos graduandos. É válido ainda indagar se o uso de ferramentas de Bioinformática de maneira ativa contribui significativamente com a aprendizagem dos alunos de graduação e consequentemente na formação destes futuros profissionais.



### **3 - OBJETIVOS**

#### **3.1 – Geral:**

Introduzir os principais conhecimentos da Bioinformática aos alunos de graduação do curso de Ciências Biológicas matriculados na disciplina de Genética Molecular, através da solução de problemas estabelecidos, sob o uso de ferramentas computacionais, de maneira a facilitar a compreensão dos assuntos abordados na disciplina.

#### **3.2 – Específicos:**

- Verificar se a Bioinformática auxilia na compreensão dos assuntos abordados em sala a partir do ponto de vista dos alunos.
- Verificar se as problemáticas apresentadas ampliam habilidades investigativas dos alunos de graduação;
- Observar a partir das perspectivas dos alunos a importância desta disciplina no currículo do curso.

## **4 - METODOLOGIA**

### **4.1 - Área de estudo**

O presente trabalho foi realizado na Universidade Federal de Sergipe onde as casuísticas foram alunos da graduação que aceitaram participar do trabalho proposto. A proposta foi feita para uma turma de Genética Molecular composta por 27 alunos.

### **4.2 - Procedimento técnico**

Este estudo foi submetido ao Comitê de Ética em Pesquisa Envolvendo Seres Humanos, através da Plataforma Brasil, obtendo aprovação no dia 11/12/2017 (CAAE: 80239717.3.0000.5546). Os participantes, de forma individual, foram esclarecidos sobre a pesquisa e voluntariamente assinaram o Termo de Consentimento Livre Esclarecido (TCLE)<sup>1</sup> anexado, seguindo o preconizado na resolução 510/2016.

Foram planejadas 20 aulas a fim de introduzir os principais aspectos da Bioinformática bem como apresentar as ferramentas utilizadas, como demonstrado na Tabela 01. Para todas as aulas que envolviam o uso das ferramentas computacionais foram elaboradas abordagens teóricas e atividades prática, onde os alunos compreendiam a teoria de cada ferramenta na prática. Para a realização dos procedimentos práticos os alunos utilizaram tutoriais e roteiros<sup>2</sup> elaborados especificamente para este fim. Desta maneira o aluno tinha a oportunidade de conhecer melhor cada ferramenta e tirar possíveis dúvidas sobre o funcionamento do programa utilizado.

Todas as aulas foram planejadas com o propósito de criar o melhor cenário possível para que os alunos, em seguida, pudessem explorar as ferramentas de maneira autônoma, discutindo em grupo e atuando de forma ativa sobre as problemáticas a eles apresentadas.

Ao todo foram elaborados quatro tutoriais e um roteiro, sendo os tutoriais utilizados para as aulas de análise de cromatogramas, alinhamento de sequências, alinhamentos múltiplos de sequências, e filogenia molecular, e o roteiro foi elaborado para aplicação na aula de banco de dados.

---

<sup>1</sup> Em apêndice: Termo de Consentimento Livre Esclarecido (TCLE)

<sup>2</sup> Em apêndice: Tutoriais e Roteiro.

Tabela 01: Tabela com os temas das aulas, horas /aula dedicada a cada tema.

ATIVIDADE	HORA/AULA
Introdução à Bioinformática	2
Análise de cromatogramas	2
Banco de Dados	2
Alinhamento de sequências	4
Filogenia Molecular	4
Dúvidas	4
Apresentação das soluções dos problemas	2

Além das aulas teórico-práticas, quatro aulas ficaram disponíveis para que os alunos sanassem dúvidas relacionadas ao funcionamento dos softwares que implicassem na resolução do seu problema.

As aulas teóricas buscavam abordar as principais características do tema em questão além de demonstrar o funcionamento dos programas que seriam utilizados. Após as aulas, os alunos resolviam em sala um problema de baixa complexidade com o auxílio de um tutorial e do professor, desta maneira os alunos se familiarizavam com os programas utilizados.

Após todas as aulas teóricas e práticas, os alunos formaram 5 grupos e participaram de um sorteio que definiu seus temas e problemas biológicos a serem resolvidos.

Cada problemática teria algumas questões que deveriam ser respondidas e os alunos deveriam preparar uma apresentação onde demonstrariam suas problemáticas, seus resultados e as metodologias utilizadas.

Cada grupo ficou responsável por encontrar soluções para as problemáticas a eles apresentadas da maneira que acreditassem ser a melhor possível.

Observa-se que o propósito em estabelecer problemas é de estimular o aluno em buscar as soluções possíveis para tais de maneira autônoma, visando ampliar o pensamento crítico, o raciocínio e habilidades investigativas dos alunos.

Cada grupo ficou com o tema de uma das aulas, e suas respectivas problemáticas, como demonstrado na Tabela 02.

Tabela 02: Distribuição dos temas para os grupos formados.

Grupo	Tema
Grupo 01	Análise de cromatogramas
Grupo 02	Banco de dados
Grupo 03	Alinhamento com o BLAST
Grupo 04	Alinhamento múltiplo
Grupo 05	Filogenia

As problemáticas de cada grupo foram as seguintes:

#### **Problemática 01: Análise de sequenciamento**

Vocês fazem parte de uma equipe de laboratório de bioinformática, e receberam de um outro laboratório alguns resultados de um sequenciamento do tipo Sanger que foi realizado. O referido laboratório solicita seus serviços a fim de responder as seguintes questões:

- Todas as amostras apresentaram um bom sequenciamento? Justifique.
- Qual o valor de corte de PHRED utilizado nesta análise? Por que?
- Os contigs foram formados da maneira correta? Qual o tamanho obtido na sequência consenso?

#### **Problemática 02: Bancos de Dados**

Seu orientador está solicitando um conjunto de informações a respeito do gene E6 de Papiloma vírus humano (HPV) e solicita as seguintes respostas:

- Qual banco de dados você irá utilizar para acessar as informações de um único gene? E porquê?
- Qual o tamanho do gene, localização do gene no genoma do organismo e quantos aminoácidos tem a proteína?
- Qual a função/funções do gene em questão?

#### **Problemática 03: Alinhamento com o BLAST**

Um grupo de pesquisa dividiu as tarefas para a execução de um trabalho, você e sua equipe ficaram responsáveis por identificar a que organismos pertence um conjunto de sequências, encontrar espécies geneticamente similares às amostras que lhe foram encaminhadas através de uma ferramenta de busca por alinhamento local. Em seguida solicitaram as seguintes respostas:

- Existem 5 formas de realizar buscas de acordo com as sequências que você obtém, sabendo disto, qual será o método que você irá utilizar (BLASTn, BLASTp, tBLASTn...)? Justifique.
- Os resultados encontrados na sua busca são estatisticamente significantes? Por quê?

#### **Problemática 04: Alinhamento múltiplo de sequências**

Ao receber um conjunto de sequências, seu orientador solicita que você realize um alinhamento e faça uma análise simples sobre estas, e pede que você verifique:

- a) Qual método você escolheria para realizar o alinhamento? ClustalW ou MUSCLE? Por que motivo?
- b) Após o alinhamento foi possível verificar que no 4º sítio há uma substituição de bases, esta substituição muda o aminoácido correspondente? Como denominamos este tipo de mutação?
- c) No sítio 634 é possível observar que um *Gap* foi inserido na espécie *Tarsius syrichta*, esse *Gap* modificou a matriz de leitura da sequência? Se sim, como minimizar os efeitos da inserção desse *Gap* levando em consideração a matriz de leitura?

#### **Problemática 05: Filogenia molecular**

Dentre as tarefas divididas no seu grupo de pesquisa, você e sua equipe receberam a função de realizar a filogenia de alguns genes alinhados. Para realizar a construção filogenética seu chefe lhe pergunta o seguinte:

- a) Dentre os métodos de construção de árvores filogenéticas, qual irá utilizar e por que motivo? (Neighbor-joining, máxima verossimilhança, máxima parcimônia)
- b) Esta árvore é confiável? Ela consegue explicar as relações filogenética entre as espécies analisadas? Por que?

### **4.3 - Coleta de dados**

A coleta de dados foi efetuada através de um questionário<sup>3</sup> de 11 questões, em que oito foram questões subjetivas, visando obter maior clareza nas respostas, e três foram objetivas, facilitando a categorização dos dados. Os dados obtidos foram tabulados (para questões objetivas) e analisados individualmente (para questões subjetivas). O questionário foi validado com três estudantes de graduação que não participaram do grupo de estudo para este projeto.

---

<sup>3</sup> Em apêndice: Questionário aplicado.

## 5 - RESULTADOS E DISCUSSÃO

Foram recebidos 23 questionários respondidos. Dentre estes, 47% (11 participantes) responderam à primeira questão de maneira semelhante, destacando a importância da introdução da Bioinformática na graduação, além da falta de informação acerca deste tema, isso pode ser observado em algumas das respostas abaixo:

Aluno 01: “Importante. Porque temos pouco contato com esse tipo de assunto, ainda mais quanto as práticas.”

Aluno 02: “É interessante, porque é algo que não temos no currículo e é uma área muito utilizada na área de pesquisas e laboratorial.”

Aluno 03: “Achei importante porque introduziu conhecimentos além da grade curricular, abrindo novas perspectivas para outras formas de emprego na área da Biologia.”

Aluno 04: “Achei muito interessante, pois é uma área pouco falada nas disciplinas de graduação e atualmente muito usada na pós-graduação [...]”

Aluno 05: “De grande proveito, pois não temos essa disciplina no currículo do nosso curso.”

Com base nestas citações é notável que os alunos percebem que existe a necessidade de utilizar a Bioinformática ainda na graduação, mesmo que de maneira mais superficial, pois, além de trazer aspectos que auxiliam na aprendizagem destes alunos, os conhecimentos básicos sobre esta disciplina provavelmente serão muito importantes para as pesquisas de pós-graduação de muitos destes alunos.

De acordo com Wightman e Hark (2012), a introdução de conceitos de Bioinformática aos alunos no início de suas carreiras é eficaz e resulta em ganhos de aprendizado significativos. O que foi bastante evidenciado nas respostas dos próprios alunos.

Os demais alunos resumiram suas ideias referente à esta questão alegando que a introdução foi bastante útil para uma melhor compreensão das análises feitas em laboratório.

A segunda questão indagava aos participantes se as aulas de Bioinformática ampliavam suas habilidades investigativas. 82% (19 participantes) disseram acreditar que suas habilidades investigativas foram de fato ampliadas de acordo com a abordagem do conteúdo. Já a avaliação feita pelos alunos sobre o curso, de maneira geral, foi bastante positiva, em que 73% (17 participantes) avaliaram como “ótimo” e o restante como “bom”.

Essas duas questões conseguem demonstrar que a introdução desta disciplina conseguiu contribuir com aspectos importantíssimos para o futuro profissional. Estabelecer o pensamento crítico e analítico não tem sido uma tarefa fácil na docência e o uso destas ferramentas sob esta abordagem podem contribuir com este propósito.

Sobre a importância e relevância da Bioinformática no currículo da graduação em Biologia, 22 participantes consideraram importante para a sua formação, além disto, todos os alunos concordaram que de fato esta disciplina é relevante para o currículo.

Porém, alguns atribuíram maior importância desta disciplina para auxiliar na aprendizagem de outras disciplinas como por exemplo Genética e Evolução., enquanto outros (56%) disseram ampliar bastante a área de atuação dos formandos e prepara-los mais para o mercado de trabalho.

Um dos pontos fundamentais para este trabalho trata-se da metodologia utilizada para abordar os temas escolhidos, pois utilizou-se das práxis na busca de atribuir significado aos conceitos fundamentais da Bioinformática. Os participantes da pesquisa também foram convocados à avaliarem este quesito através das questões 6 e 7.

Com relação a este quesito, 100% dos participantes consideraram a metodologia ativa mais eficiente do que a abordagem apenas expositiva, alguns até justificam o motivo pelo qual esta metodologia foi bem avaliada por eles, como pode ser observado a seguir:

Aluno 01: “É mais eficiente, uma vez que torna o aluno “protagonista”, ele quem vai buscar e construir o conhecimento dentro dos assuntos que são expostos.”

Aluno 04: “Muito eficiente, pois estimula o aluno.”

Aluno 06: “Mais eficiente, pois, fez a gente exercitar o que aprendemos na teoria e assim saber identificar em que tipo de situação podemos aplicá-la na prática.”

Aluno 07: “Muito mais efetiva, pois se somente fosse trabalhado a exposição sem a busca ativa, veríamos somente problemáticas fechadas e que não representariam problemas reais que pudéssemos enfrentar em nossas carreiras. ”

A investigação estimula a criatividade e curiosidade, e incrementa a atitude crítica dos alunos, que desenvolvem uma nova postura em relação ao saber e se tornam mais motivados para aprender (Rodrigues e Borges, 2008; Munford e Lima, 2007; Zompero e Laburú, 2010).

Em relação às sugestões para a abordagem da disciplina, alguns alunos trouxeram características interessantes que devem ser levadas em consideração, por exemplo: Aluno 01: “seguir um passo a passo desde a coleta do material até as análises”, ou ainda, “cada aluno pensar em sua problemática e buscar maneiras de solucioná-la”.

Estas falas conotam o quanto os alunos ficam entusiasmados ao sentir-se parte do processo de ensino-aprendizagem, bem como se sentem mais protagonistas ao participar ativamente da construção do seu conhecimento.

Quanto a abordagem exclusiva da disciplina, os participantes dividiram opiniões, alguns acreditam que com uma abordagem exclusiva haveria a possibilidade de maior aprofundamento dos conteúdos além de mais tempo para a realização das atividades. Outros, acreditam que a interdisciplinaridade pode ser a chave para uma abordagem eficiente, uma vez que a Bioinformática é utilizada em diversas áreas.

Alguns alunos mencionaram que de fato a abordagem interdisciplinar facilita o processo de aprendizagem, como visto nas falas a seguir:

Aluno 01: “Poderia ser abordada com a disciplina de Genética Molecular, já que os assuntos estão interligados e ela ajudou no entendimento da genética”

Aluno 04: “É importante ser incluída em outras disciplinas, para fixar o conteúdo.  
”

Aluno 05: “Em conjunto, porque várias outras disciplinas poderiam usar artifícios para melhorar o entendimento dos assuntos. ”

Aluno 12: “Em conjunto com outras disciplinas, porque fica mais interativa. ”

Aluno 13: “Acho que poderia haver uma interdisciplinaridade, pois acaba ajudando aqueles que não possuem facilidade em genética. ”



Aluno 15: “Em conjunto, pois pode auxiliar no aprendizado de outras matérias”

Aluno 17: “Ela deveria ser abordada sim com outras matérias, afinal o conhecimento é integrado e quanto mais essa integração é mostrada mais facilmente se constroem links nas informações técnicas aprendidas. ”

Aluno 19: “Em conjunto com a genética, uma vez que estão interligadas. ”

Aluno 20: “Em conjunto, pois aproxima os alunos da realidade da pesquisa, que é interdisciplinar. ”

De acordo com as falas acima, é possível notar que o uso da Bioinformática através da metodologia ativa, consegue estimular e facilitar a aprendizagem dos alunos para os temas abordados. Observa-se também que as aulas de Bioinformática foram capazes de auxiliar no ensino de Genética. Os alunos ainda citam exemplos de outras disciplinas que poderiam usar destas ferramentas para ajudar na compreensão como Evolução e Bioquímica.

Com relação aos conhecimentos prévios sobre Bioinformática, 12 participantes alegaram não conhecer a disciplina antes desta abordagem, os demais disseram conhecer superficialmente o assunto. O que reforça a falta de informação desta área em ascensão aos alunos de graduação.

## 6 - CONCLUSÕES

Com subsídio no exposto, é perceptível que poucos alunos conhecem a Bioinformática, e aqueles que conhecem, muitas vezes não sabem como utilizar as suas ferramentas nem em quais áreas se aplicam. Desta forma, é notória a importância da introdução das ferramentas de Bioinformática no currículo da graduação em Ciências Biológicas, uma vez que ela se mostrou facilitadora do processo de ensino-aprendizagem dos alunos, além de ampliar habilidades consideradas fundamentais para estes profissionais e sobretudo estimular os alunos a elaborarem pensamentos críticos sobre os assuntos explorados.

Os participantes da pesquisa conseguiram identificar que de fato esta disciplina os auxiliou na sua percepção de possíveis soluções para uma problemática, e, portanto, trazer a Bioinformática para a graduação é de extrema importância para a formação acadêmica dos graduandos, gerando recursos humanos capacitados não apenas para o mercado de trabalho imediato como também para a investidura na pós-graduação.

A metodologia ativa empregada facilitou a aprendizagem dos conteúdos de genética vistos anteriormente, fixando o conteúdo e auxiliando nas construções de significados para os assuntos estudados.

Sendo assim é possível inferir que a aplicação de atividades como as propostas nesse trabalho agrega conhecimentos, ampliam habilidades importantes para o futuro profissional, e facilitam a aprendizagem de temas abstratos como a genética por exemplo, isto demonstra como estas metodologias podem contribuir para melhorar a qualidade da formação do biólogo.

## REFERÊNCIAS

- ALVES, S. M. A bioinformática e sua importância para a biologia molecular. **Revista Brasileira de Educação e Saúde**, v. 3, n. 4, p. 18–25, 2013.
- FARAH, S. B. **DNA segredos e mistérios**. Sarvier ed. São Paulo: [s.n.].
- GENBANK. **GenBank and WGS Statistics**. Disponível em: <<https://www.ncbi.nlm.nih.gov/genbank/statistics/>>. Acesso em: 1 set. 2017.
- GEWIN, V. Education: Inspiration for informatics. **Nature**, v. 478, n. 7367, p. 143–145, 2011.
- GOLD. **Statistics Complete and Permanent Draft Genome Totals in GOLD (by year and status)**. Disponível em: <<https://gold.jgi.doe.gov/statistics>>. Acesso em: 1 set. 2017.
- KOCH, I.; FUELLEN, G. A review of bioinformatics education in Germany. **Briefings in Bioinformatics**, v. 9, n. 3, p. 232–242, 11 jan. 2008.
- MELO, B.; SANT’ANA, G. A prática da Metodologia Ativa: compreensão dos discentes enquanto autores do processo ensino- aprendizagem. **Esics.Edu.Br**, v. 23, n. 4, p. 327–339, 2013.
- MOREIRA, L. C.; SOUZA, G. S. DE; ALMASSY, R. C. B. AS ATIVIDADES INVESTIGATIVAS E A RESOLUÇÃO DE PROBLEMAS NO ENSINO DE BIOLOGIA: LIMITES E POSSIBILIDADES. **EREBIO**, p. 2687–2698, 2014.
- NUNES, R.; BARBOSA DE ALMEIDA JÚNIOR, E.; PESSOA PINTO DE MENEZES, I.; MALAFAIA, G. Learning nucleic acids solving by bioinformatics problems. **Biochemistry and Molecular Biology Education**, v. 43, n. 5, p. 377–383, 2015.
- OUZOUNIS, C. A. Rise and demise of bioinformatics? promise and progress. **PLoS Computational Biology**, v. 8, n. 4, p. 1–5, 2012.
- PEREZLEO, LI.; RICARDO, S.; JORGE, A.; CONILL, C.; GUDELIA, G.; VELOZ, A.; RUIZ, A. A. Impacto de la Bioinformática en las ciencias biomédicas. **Acimed**, v. 11, n. 4, p. 1–6, 2003.
- RIBEIRO, D. DA S.; JESUS, T. F. DE. A Práxis Pedagógica no Ensino Superior: um relato baseado na necessidade de ações inovadoras de avaliação. **Cairu em Revista**, v. 4, n. Jan/Fev nº05, p. 124–142, 2015.
- RIBEIRO JUNIOR, H. L.; OLIVEIRA, R. T. G. DE; CECCATTO, V. M. Bioinformática como recurso pedagógico para o curso de ciências biológicas na Universidade Estadual do Ceará – UECE – Fortaleza, Estado do Ceará. **Acta Scientiarum. Education**, v. 34, n. 1, p. 129–140, 2012.
- VENTER, J. C. Multiple personal genomes await. **Nature**, v. 464, n. April, p. 676–677, 2010.
- WIGHTMAN, B.; HARK, A. T. Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. **Biochemistry and Molecular Biology Education**, v. 40, n. 5, p. 310–319, 2012.

ZATZ, M. M. Bioinformatics training in the USA. **Briefings in bioinformatics**, v. 3, n. 4, p. 353–60, 2002.

ZETI, A. M. H.; SHAMSIR, M. S.; TAJUL-ARIFIN, K.; MERICAN, A. F.; MOHAMED, R.; NATHAN, S.; MAHADI, N. M.; NAPIS, S.; TAN, T. W. Bioinformatics in Malaysia: Hope, initiative, effort, reality, and challenges. **PLoS Computational Biology**, v. 5, n. 8, p. 2007–2010, 2009.

ALVES, S. M. A bioinformática e sua importância para a biologia molecular. **Revista Brasileira de Educação e Saúde**, v. 3, n. 4, p. 18–25, 2013.

FARAH, S. B. **DNA segredos e mistérios**. Sarvier ed. São Paulo: [s.n.].

GENBANK. **GenBank and WGS Statistics**. Disponível em: <<https://www.ncbi.nlm.nih.gov/genbank/statistics/>>. Acesso em: 1 set. 2017.

GEWIN, V. Education: Inspiration for informatics. **Nature**, v. 478, n. 7367, p. 143–145, 2011.

GOLD. **Statistics Complete and Permanent Draft Genome Totals in GOLD (by year and status)**. Disponível em: <<https://gold.jgi.doe.gov/statistics>>. Acesso em: 1 set. 2017.

KOCH, I.; FUELLEN, G. A review of bioinformatics education in Germany. **Briefings in Bioinformatics**, v. 9, n. 3, p. 232–242, 11 jan. 2008.

MELO, B.; SANT'ANA, G. A prática da Metodologia Ativa: compreensão dos discentes enquanto autores do processo ensino- aprendizagem. **Esics.Edu.Br**, v. 23, n. 4, p. 327–339, 2013.

MOREIRA, L. C.; SOUZA, G. S. DE; ALMASSY, R. C. B. AS ATIVIDADES INVESTIGATIVAS E A RESOLUÇÃO DE PROBLEMAS NO ENSINO DE BIOLOGIA: LIMITES E POSSIBILIDADES. **EREBIO**, p. 2687–2698, 2014.

NUNES, R.; BARBOSA DE ALMEIDA JÚNIOR, E.; PESSOA PINTO DE MENEZES, I.; MALAFAIA, G. Learning nucleic acids solving by bioinformatics problems. **Biochemistry and Molecular Biology Education**, v. 43, n. 5, p. 377–383, 2015.

OUZOUNIS, C. A. Rise and demise of bioinformatics? promise and progress. **PLoS Computational Biology**, v. 8, n. 4, p. 1–5, 2012.

PEREZLEO, LI.; RICARDO, S.; JORGE, A.; CONILL, C.; GUDELIA, G.; VELOZ, A.; RUIZ, A. A. Impacto de la Bioinformática en las ciencias biomédicas. **Acimed**, v. 11, n. 4, p. 1–6, 2003.

RIBEIRO, D. DA S.; JESUS, T. F. DE. A Práxis Pedagógica no Ensino Superior: um relato baseado na necessidade de ações inovadoras de avaliação. **Cairu em Revista**, v. 4, n. Jan/Fev nº05, p. 124–142, 2015.

RIBEIRO JUNIOR, H. L.; OLIVEIRA, R. T. G. DE; CECCATTO, V. M. Bioinformática como recurso pedagógico para o curso de ciências biológicas na Universidade Estadual do Ceará – UECE – Fortaleza, Estado do Ceará. **Acta Scientiarum. Education**, v. 34, n. 1, p. 129–140, 2012.

VENTER, J. C. Multiple personal genomes await. **Nature**, v. 464, n. April, p. 676–677, 2010.

WIGHTMAN, B.; HARK, A. T. Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills.

**Biochemistry and Molecular Biology Education**, v. 40, n. 5, p. 310–319, 2012.

ZATZ, M. M. Bioinformatics training in the USA. **Briefings in bioinformatics**, v. 3, n. 4, p. 353–60, 2002.

ZETI, A. M. H.; SHAMSIR, M. S.; TAJUL-ARIFIN, K.; MERICAN, A. F.; MOHAMED, R.; NATHAN, S.; MAHADI, N. M.; NAPIS, S.; TAN, T. W.

Bioinformatics in Malaysia: Hope, initiative, effort, reality, and challenges. **PLoS Computational Biology**, v. 5, n. 8, p. 2007–2010, 2009.

## **APÊNDICE**



**UNIVERSIDADE FEDERAL DE SERGIPE  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE – CCBS  
DEPARTAMENTO DE BIOLOGIA - DBI**

**Termo de consentimento Livre e Esclarecido**

**INSERÇÃO DA BIOINFORMÁTICA NO ENSINO DE GENÉTICA PARA OS  
CURSOS DE GRADUAÇÃO DE CIÊNCIAS BIOLÓGICAS SOB  
METODOLOGIA ATIVA**

Prezado (a):

Gostaríamos de convidá-lo (a) para participar da pesquisa **“INSERÇÃO DA BIOINFORMÁTICA NO ENSINO DE GENÉTICA PARA OS CURSOS DE GRADUAÇÃO DE CIÊNCIAS BIOLÓGICAS SOB METODOLOGIA ATIVA”** a ser realizada nesta instituição de ensino. O objetivo da pesquisa é apresentar a disciplina de bioinformática aos alunos de graduação do curso de ciências biológicas através da solução de problemas estabelecidos, sob o uso de ferramentas de bioinformática. Sua participação é muito importante e ela se daria da seguinte forma (Responder ao questionário que lhe será entregue buscando relatar as respostas de maneira mais fidedigna possível).

Esclarecemos que sua participação é totalmente voluntária, podendo o (a) senhor (a): recusar-se a participar, ou mesmo desistir a qualquer momento, sem que isto acarrete qualquer ônus ou prejuízo à sua pessoa. Esclarecemos, também, que suas informações serão utilizadas **somente** para os fins desta ou futuras pesquisas e serão tratadas com o **mais absoluto sigilo e confidencialidade**, de modo a preservar a sua identidade.

Esclarecemos ainda, que o (a) senhor (a) não pagará e nem será remunerado (a) por sua participação. Garantimos, no entanto, que todas as despesas decorrentes da pesquisa serão ressarcidas, quando devidas e decorrentes especificamente de sua participação.

Os benefícios esperados são a contribuição no acréscimo de informações relacionadas ao ensino da bioinformática. Quanto aos riscos, não há nenhum risco na participação desta pesquisa.

Caso o (a) senhor (a) tenha dúvidas ou necessite de maiores esclarecimentos poderá nos contatar (**Everton Silva Mota – [evertonmotabio@hotmail.com](mailto:evertonmotabio@hotmail.com) – (79) 9 9139-4167 ou (79) 9 9119-5776**

Este termo deverá ser preenchido em duas vias de igual teor, sendo uma delas devidamente preenchida, assinada e entregue ao (à) senhor (a).

Sergipe, \_\_\_\_\_ de \_\_\_\_\_ de 201\_\_\_\_.

**Pesquisador Responsável RG: \_\_\_\_3.426.794 - 8\_\_\_\_**



Eu, \_\_\_\_\_, tendo sido devidamente esclarecido sobre os procedimentos da pesquisa, concordo em participar **voluntariamente** da pesquisa descrita acima.

Assinatura: \_\_\_\_\_

Data: \_\_\_\_\_





## Questionário

- 1) O que você achou sobre a introdução da Bioinformática para a turma de graduação?  

---

---

---

---

---
- 2) Você acredita que as aulas de Bioinformática ampliaram as suas habilidades investigativas?  
  
( ) sim ( ) um pouco ( ) não
- 3) Como você avaliaria a inserção da Bioinformática feita para sua turma?  
  
( ) Ótimo ( ) Bom ( ) Regular ( ) Ruim
- 4) Na sua opinião qual a relevância da Bioinformática no currículo da graduação em Biologia?  

---

---

---

---

---
- 5) Sobre a importância da Bioinformática na sua formação acadêmica, como você a classificaria?  
  
( ) Muito importante  
( ) Importante  
( ) Pouco importante  
( ) Sem importância
- 6) Você acha que apresentar a disciplina sob uma abordagem ativa, onde o aluno soluciona problemas foi mais eficiente ou menos eficiente do que uma abordagem apenas expositiva?  

---

---

---

---

---



- 7) Na sua opinião, apenas as aulas teóricas seriam suficientes para que os alunos entendessem a bioinformática? Por que?

---

---

---

---

---

---

---

- 8) Há alguma sugestão referente à abordagem da Bioinformática?

---

---

---

---

---

---

---

- 9) Você acredita que a Bioinformática deveria ser abordada de maneira exclusiva ou em conjunto com outras disciplinas? Por que?

---

---

---

---

---

- 10) Você já tinha algum conhecimento sobre a Bioinformática antes desta disciplina? Caso sim, como você conheceu esta área?

---

---

---

---

- 11) Você utilizará (ia) a Bioinformática como uma ferramenta para os estudos e/ou trabalhos científicos? Se sim, porque a escolheria? Caso a resposta seja negativa, o que lhe impede (iria) de usá-la?

---

---

---

---

---



## Tutorial Staden

**Observações:** Sugerimos nomear as pastas dos arquivos não colocando espaços, acentos ou símbolos, pois estes podem acabar influenciando no funcionamento do programa.

É sugerido que as pastas contendo os arquivos sejam armazenadas no disco C do computador, evitando que o programa não reconheça os arquivos, impedindo a abertura dos mesmos.

### Introdução:

O Staden é um pacote de programas gratuito que roda em Windows, Unix, Linux e MacOSX, cuja finalidade é tornar uma série de sequências sobrepostas curtas em uma grande sequência contígua. O pacote de programas fornece ferramentas para analisar essas sequências.

Staden está disponível em <http://staden.sourceforge.net/>. Para utilizá-lo você precisa fazer o download da versão compatível com o seu computador, instalá-lo.



The screenshot shows the Staden SourceForge project page. The browser window has tabs for 'MSN Brasil - Outlook, Hotmail' and 'Staden Package Inicio'. The address bar shows 'staden.sourceforge.net'. The page has a green header with the title 'Pacote Staden'. Below the header is a navigation bar with links: 'Casa', 'SF Inicio', 'De downloads' (highlighted), 'Imagens', 'Documentação', and 'Fóruns'. The main content area is divided into four sections: 'Introdução', 'Notícia', 'Lançamentos recentes', and 'Documentação'. The 'Introdução' section contains a description of the software and links to 'Casa projeto SourceForge', 'Resumo', and 'Quem é quem'. The 'Lançamentos recentes' section lists several releases from December 2013, including 'Staden-2.0.0b10-windows-x86\_64.exe' and 'Staden-2.0.0b10-linux-i386.tar.gz'. The 'Documentação' section contains links to 'Notas de lançamento', 'Manual Mini', 'Manual do usuário', 'Métodos de detecção de mutação', and 'Integração Phrap e GAP4'.



sourceforge.net/projects/staden/files/staden/2.0.0b10/

Trazido a você por: [awhitwham](#), [jbonfield](#)

Resumo Arquivos Comentários Apolo Wiki Hospedado Apps Bilhetes Notícia Discussão Código

Olhando para a versão mais recente? [Baixe staden-2.0.0b9-src.tar.gz \(4.4 MB\)](#)

Início / staden / 2.0.0b10

Nome	Modificado	Tamanho	Download / Semana
<b>Pasta pai</b>			
<a href="#">Staden-2.0.0b10-windows-x86_64.exe</a>	2013/12/18	17.7 MB	74
Clique para baixar staden-2.0.0b10-windows-x86_64.exe			
<a href="#">Staden-2.0.0b10-windows-x86.exe</a>			
<a href="#">Staden-2.0.0b10-linux-i386.tar.xz</a>			
Texto original			
<a href="#">Staden-2.0.0b10-linux-i386.tar.gz</a>			
staden-2.0.0b10-windows-x86_64.exe			
<a href="#">Staden-2.0.0b10-linux-x86_64.tar.xz</a>			
Sugira uma tradução melhor			
<a href="#">Staden-2.0.0b10-linux-x86_64.tar.gz</a>	2013/12/18	23.1 MB	19
<a href="#">Staden-2.0.0b10-MacOSX.dmg</a>	2013/12/18	25.9 MB	17
<a href="#">Staden-2.0.0b10-src.tar.gz</a>	2013/12/18	4.6 MB	11
Totais: 8 Itens			
		144,6 MB	148

sourceforge.net/projects/staden/files/staden/2.0.0b10/staden-2.0.0b10-windows-x86\_64.exe/download

**Go Parallel**  
Translating Multicore Power into Application Performance

Stay connected, up-to-date, and informed on all things parallel development via Go Parallel, where you'll find viewpoints, how-to's, software tools, and educational information to help your software

Últimas Jobs tecnologia Desenvolvido por **Disc**

Senior Java Developer (Job Title)

Encontrar

SAP PP - Gerente  
Deloitte - Detroit, MI

Desenvolvedor SharePoint  
APN Serviços de Software, Inc - Austin, TX

Os programas principais contidos no Staden para montagem de sequências são PreGap4 e Gap4.

## PreGap4

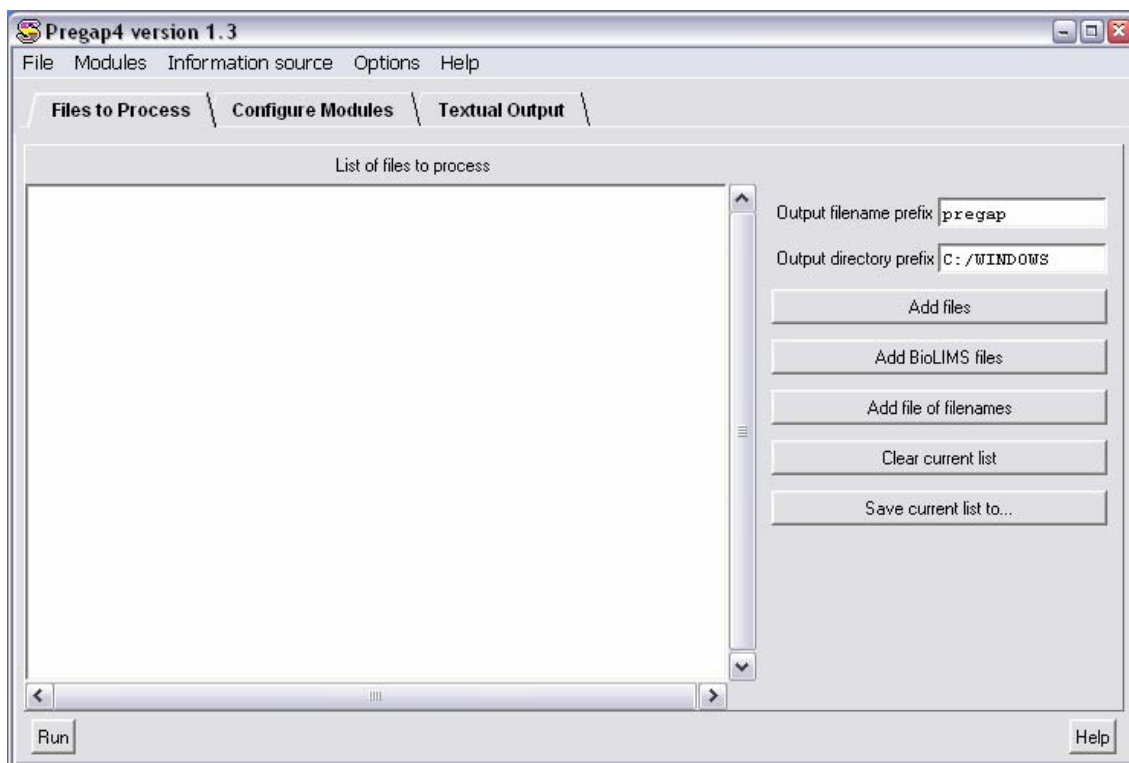
Este programa é utilizado para mascarar as sequências de má qualidade que você provavelmente não quer que apareça na montagem final. Ele vai processar arquivos “raw” para que eles estejam de forma adequada para a montagem.

## Gap4

Reúne fragmentos individuais em contigs longos permitindo a edição da sequência, ou seja, ela faz um consenso entre as sequências a fim de evitar discrepâncias.

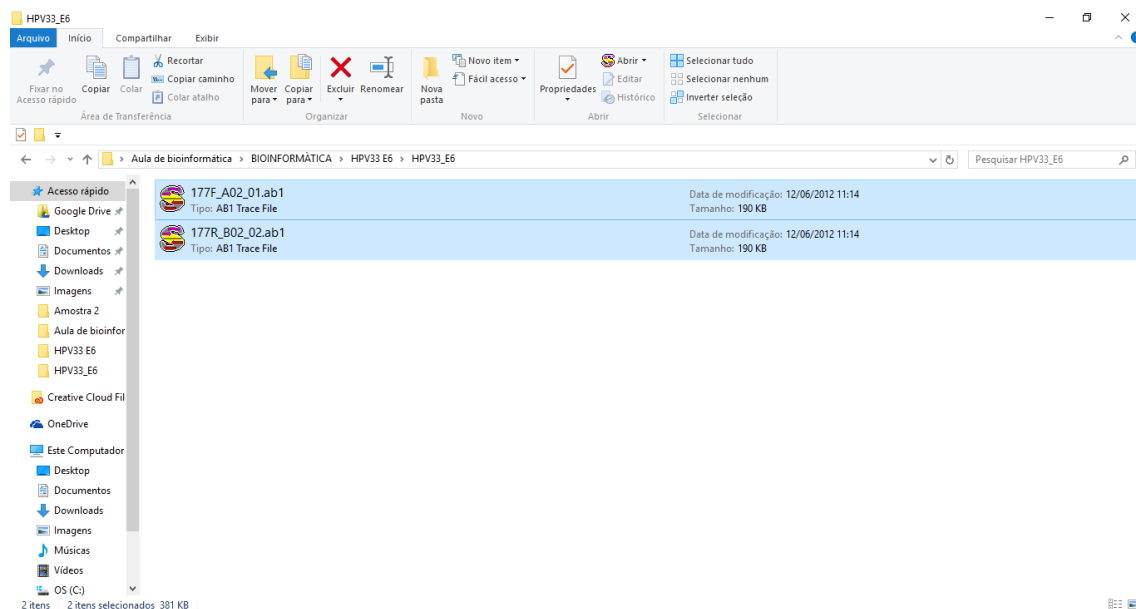
## Utilizando o PreGap4

Abrir o programa PreGap4, clicando em “Iniciar > Todos os Programas > Staden Package”. Após abrir o “PreGap4” você verá a seguinte janela contendo três abas na parte superior “Files to Process”, “Configure Modules” e “Textual Output”. Vamos começar a análise da esquerda para a direita.



## Processamento dos Arquivos

Para selecionar os arquivos de interesse clique em **“Add Files”** à direita. Este abrirá uma caixa de seleção de arquivos e você deve selecionar aqueles que se deseja trabalhar.



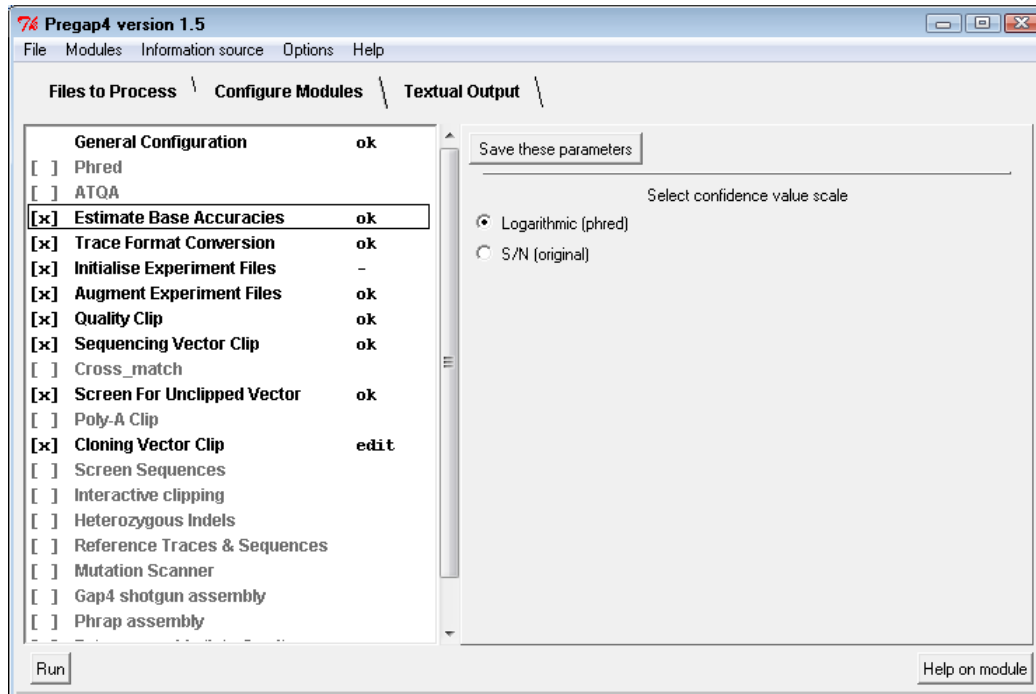
Após a seleção, os arquivos devem aparecer na janela principal do PreGap4. A próxima etapa é configurar os módulos.





## Configurar Módulos

Selecione a aba “**Configure Modules**”, onde aparecerá a seguinte tela com os módulos a esquerda da qual você pode selecionar.



Os módulos que estão em preto e marcados com o X são os módulos ativos, aqueles em cinza são os inativos. Para ativar ou desativar um módulo basta clicar no mesmo. Na maioria dos sequenciamentos não há necessidade de alterar nenhum módulo.

Veremos a definição dos módulos que devem ser marcados em uma **análise básica**:

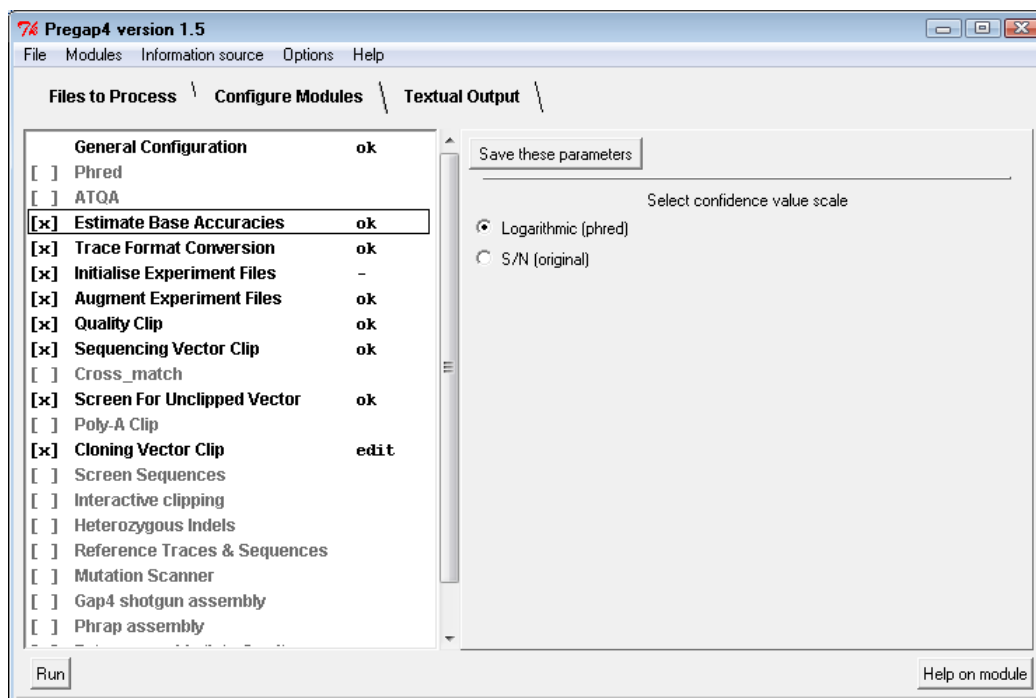
### ***General Configuration***

Nesta parte o programa pergunta se você quer olhar os seus arquivos de rastreamento para determinar o nome de cada amostra, ou se é para levar o nome do próprio arquivo. Nesta opção você deve responder não, pois a abordagem segura é ter sempre um nome a partir do arquivo.

### ***Estimate Base Accuracies***

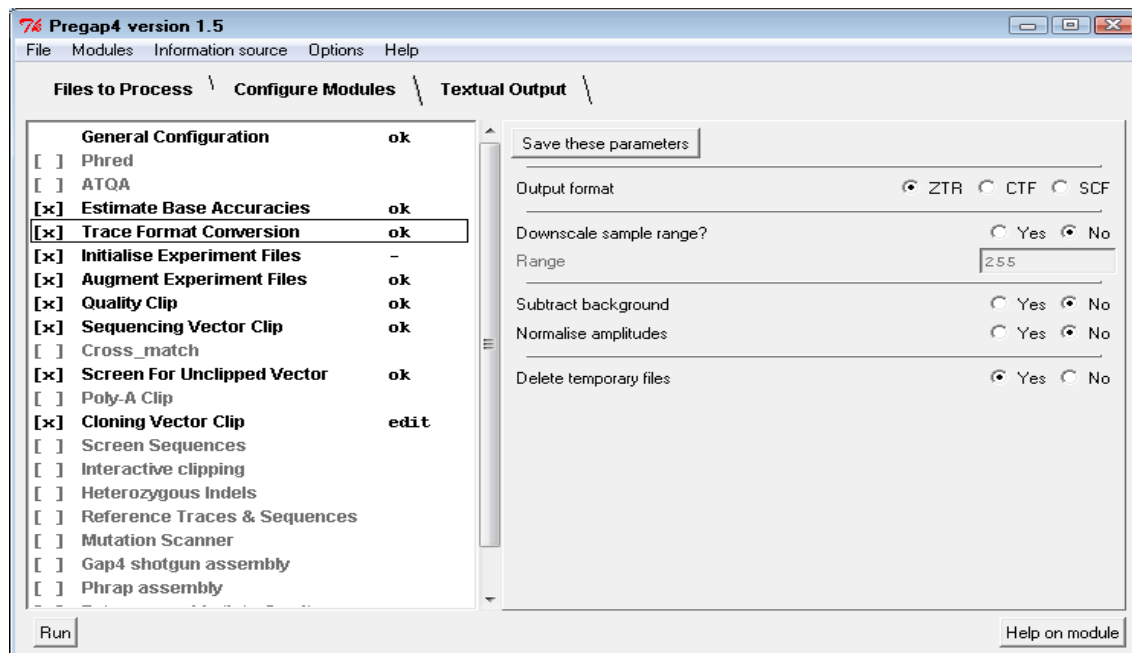
Este módulo esconde a sequência de má qualidade, ele deve ser definido como logarítmica base do cálculo do valor de Phred.





### *Trace format conversion*

Este módulo converte arquivos para um formato que Staden possa reconhecer. Convertendo para o formato ZTR você tem a vantagem de fazer arquivos muito menores o que lhe poupará espaço na hora de armazenar.



### *Initialise Experiment Files*

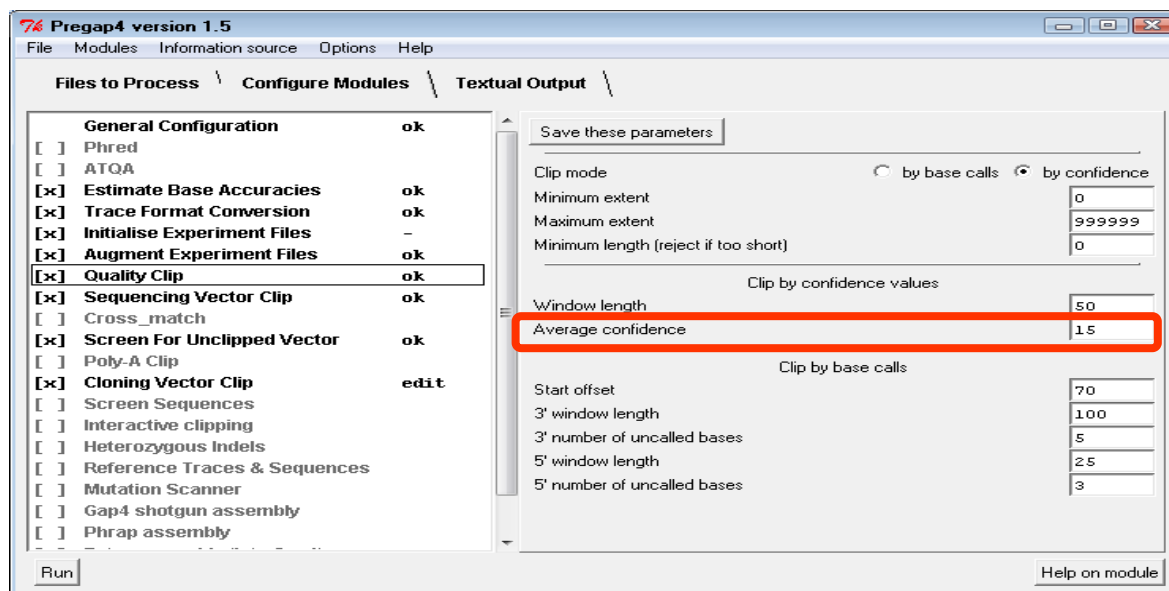
Contém arquivos de experiência que você precisará para a saída dos outros módulos. Deve-se manter esta opção ativada.





### Quality Clip

Este módulo vai mascarar sequência de má qualidade no alinhamento final, pois durante o sequenciamento, a qualidade das bases diminui no início e no fim das sequências. Existem dois modos de “clip”, um por chamada de base e outro por confiança **“By base calls”** e **“By confidence”**. Deve-se selecionar o **“by confidence”**, pois este apresenta maior precisão. Para alterar o valor de Phred, modificar a opção **“Average confidence”**.



### Sequencing Vector Clip

Este módulo pode remover a sequência de vetor.

### Interactive Clipping

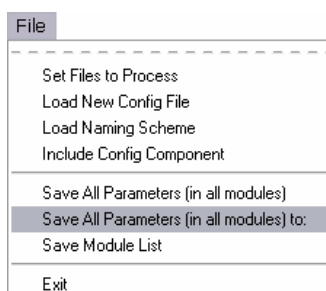
Permite a visualização dos arquivos processados.

### Cloning Vector Clip

Este módulo pode ser desativado uma vez que não executamos nenhuma clonagem.

### Salvando suas Opções

Após configurar todos os módulos você pode salvar suas configurações clicando em **“File”** no menu principal e selecionar a opção **“Save all parameters in all modules to”**, salvar com um nome qualquer e a terminação **“.aux”**.

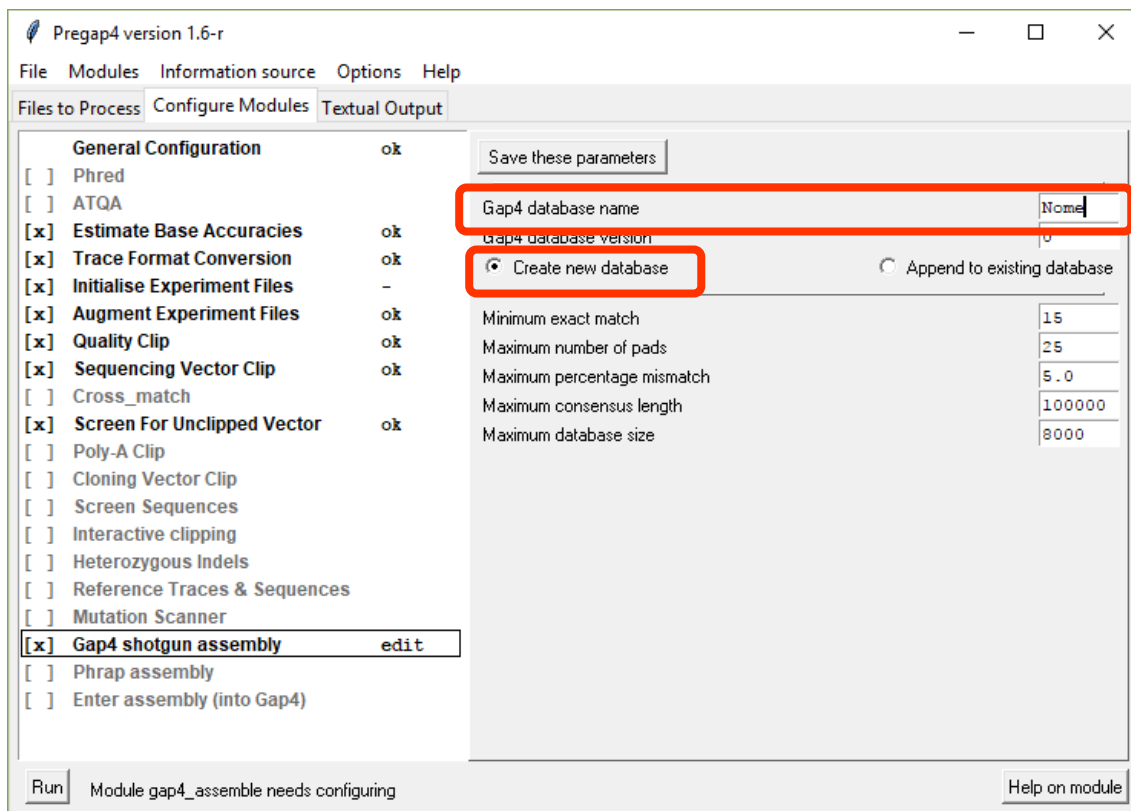






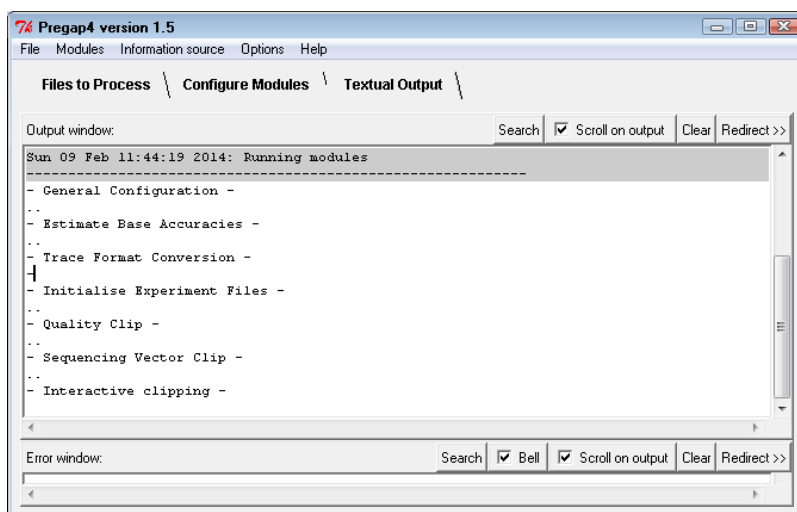
### Gap4 shotgun assembly

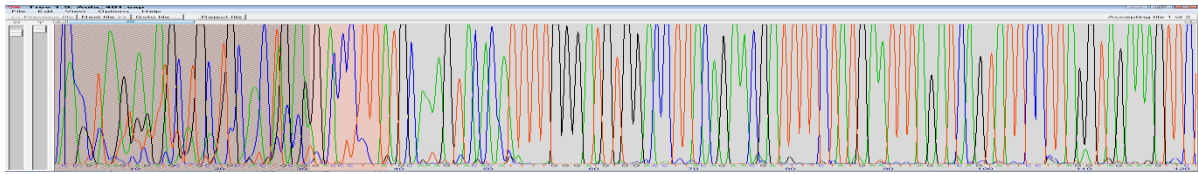
Neste módulo você irá fornecer um nome ao arquivo que será criado, para tal, deve-se clicar na aba “**create new database**” e na caixa de texto “**Gap4 database name**” deve ser colocado o nome do seu arquivo.



### Executando o PreGap4

Quando todos os módulos estiverem definidos você pode executar o programa clicando em “**Run**” na parte inferior à esquerda da janela principal. Depois de o programa ter rodado aparecerá na janela principal as seguintes informações:



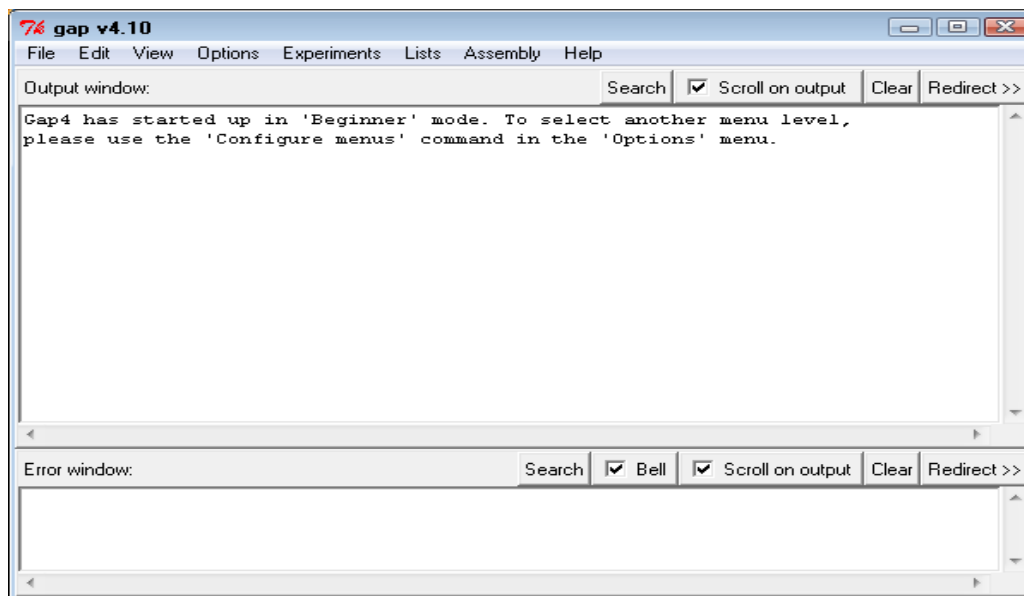


## Finalizando

Após o termino do processamento, você pode simplesmente fechar a janela.

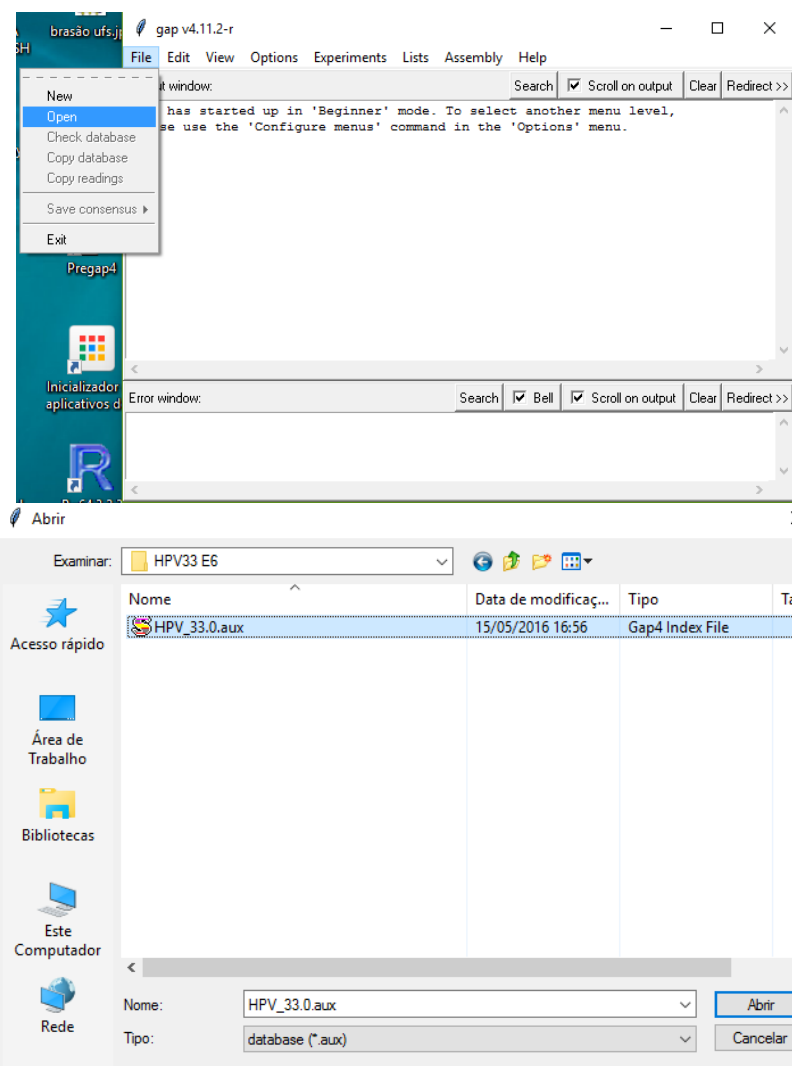
## Gap4

Após o processamento realizado pelo **PreGap4**, você pode montar a sua sequência contígua (contig) utilizando agora o Gap4. Abrindo o Gap4 (**Iniciar > Todos os Programas > Staden Package > Gap4**) você verá a seguinte janela:



## Iniciando Nova Montagem

A primeira coisa a se fazer é selecionar o arquivo com a extensão **“aux”** Isso pode ser feito apenas com dois clicks sobre o arquivo da sequência, ou clicando em (**File > Open**) e seleciona-se o arquivo desejado.



## The Contig Selector

As informações das sequências contíguas são fornecidas pelo Contig Selector nos trechos de DNA. O Contig Selector tem uma linha horizontal e uma série de linhas verticais adjacentes representando trechos de DNA montado.



## Contig options

Selecionando a linha horizontal e pressionando o botão direito do mouse aparecerão várias opções de contig, que deverão ser selecionadas:

## List notes



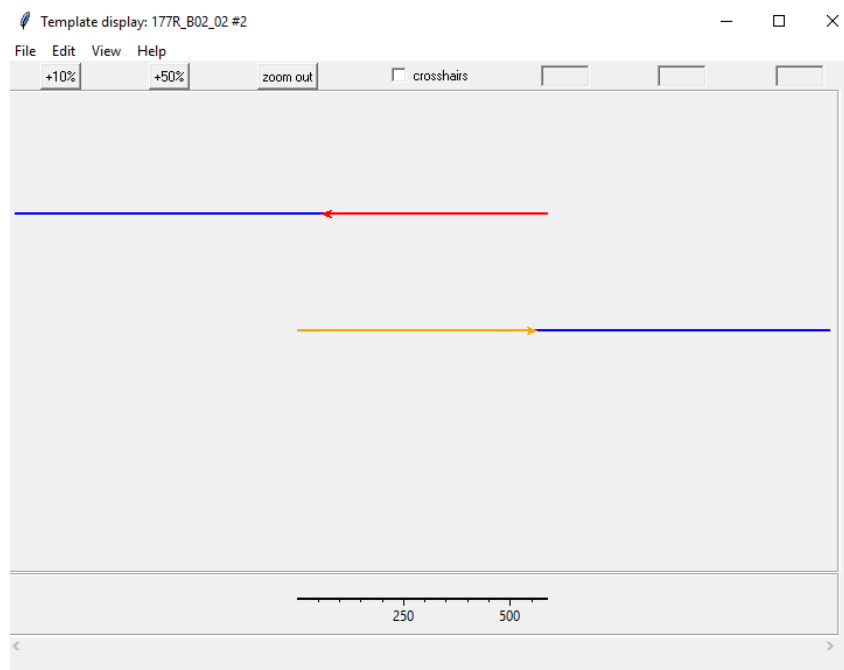
Esta opção permite anexar notas de texto para o seu contig

### Complement contig

Vão ajudar a manter a sequência boa entre os contigs curtas “sequências de grande consistência” e que os contigs longos não interfiram nos contigs curtos.

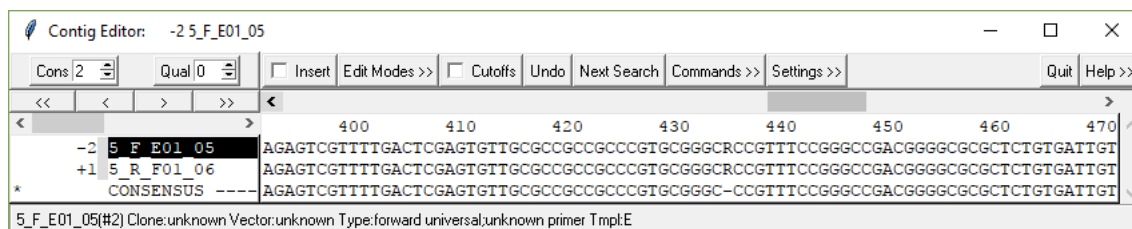
### Template display

Representação gráfica das posições e orientações de leitura dentro do contig. Você pode criar uma exibição de modelos para todos os contigs clicando em “Template display”



### The Contig Editor

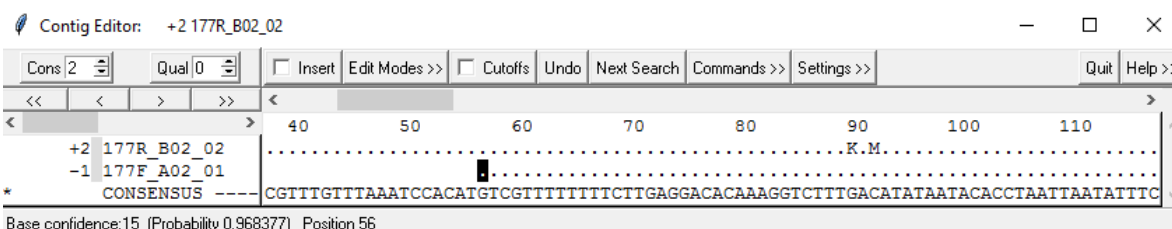
O editor contig permite que você visualize as sequências e possa identificar as diferenças entre elas. Para acessá-lo, clique com o botão direito na linha horizontal do “Contig selector” e depois clique em “Edit contig”.



O nome da sequência fica do lado esquerdo “5\_F\_E01\_05”. À esquerda do nome fica o número da sequência. Quando este é negativo, significa dizer que houve uma inversão na sequência para que ela coubesse na montagem.

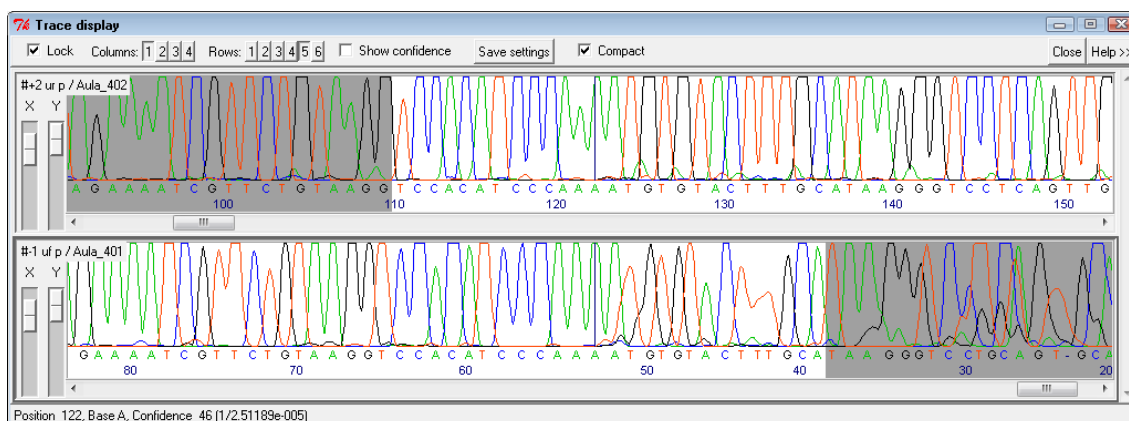
### Highlighting differences “destacando as diferenças”

Destaca diferenças entre as sequências. Clicando em “Settings” na parte superior do editor você terá as opções para encontrar as diferenças entre as sequências.



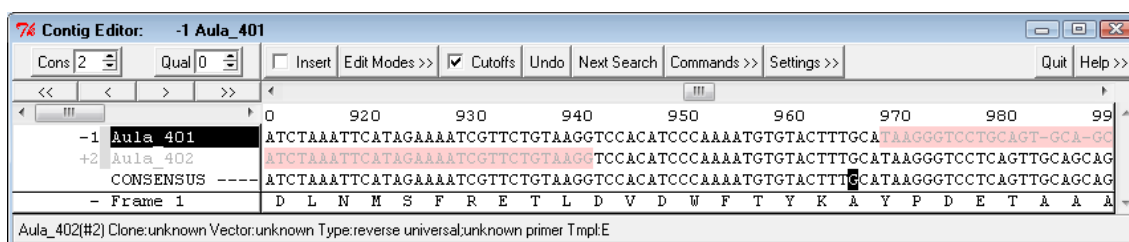
## Viewing the Chromatograms “visualizando os cromatogramas”

Você pode visualizar o cromatograma clicando duas vezes em qualquer base da sequência consenso (CONSENSUS). Para inserir os valores de Phred no cromatograma, marcar a opção “Show confidence”.



## Hidden Data

Para visualizar os dados ocultos da sequência que o PreGap considerou de má qualidade clique em “Cutoffs” na parte superior da janela “Contig Editor”.



## Salvando suas alterações

Para salvar suas alterações em contig editor basta ir à janela Gap4 clicar em “File: Save consensus: Normal” para salvar a sequência consenso (contig) no formato FASTA, porém além deste formato recomenda-se que o arquivo seja salvo em “txt” uma vez que pode ser utilizado em seguida.



## ROTEIRO:

### AULA PRÁTICA BIOINFORMÁTICA (BANCO DE DADOS)

O presente roteiro busca demonstrar maneiras de obter informações dos bancos de dados dispostos no *National Center of Biotechnology Information* (NCBI).

## Passo 1:

Acessa à página do NCBI através do link: <https://www.ncbi.nlm.nih.gov/>

Ao acessar este link a seguinte página será apresentada:

The screenshot shows the NCBI homepage with several red boxes and arrows indicating key features:

- 1:** Points to the left sidebar menu containing various resource categories like 'NCBI Home', 'Resource List (A-Z)', 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', 'Training & Tutorials', and 'Variation'.
- 2:** Points to the central 'Welcome to NCBI' section, which includes links for 'About the NCBI', 'Mission', 'Organization', and 'NCBI News & Blog', as well as icons for 'Submit', 'Download', 'Learn', 'Develop', 'Analyze', and 'Research'.
- 3:** Points to the 'Popular Resources' section on the right, listing 'PubMed', 'Bookshelf', 'PubMed Central', 'PubMed Health', 'BLAST', 'Nucleotide', 'Genome', 'SNP', 'Gene', 'Protein', and 'PubChem'.
- 4:** Points to the 'NCBI News & Blog' section, featuring a recent article about 'IgBLAST 1.8.0 release'.
- 5:** Points to the 'All Databases' dropdown menu in the top navigation bar.
- 6:** Points to the search bar in the top navigation bar.

Esta é a interface do NCBI onde você pode ver:

- 1= Lista com bancos de dados do NCBI
- 2= Opções de navegação do NCBI (depósito de sequências, download, tutoriais etc...)
- 3= Principais bancos de dados
- 4= Notícias sobre o NCBI
- 5= Caixa seletora de banco específico
- 6= Caixa de texto onde coloca-se os termos a serem pesquisados nos bancos de dados.

Para este roteiro iremos focar nos itens 5 e 6, onde escolheremos o banco de dados de nucleotídeo e executaremos uma busca para o gene *Rbcl* de *Cocos nucifera*.

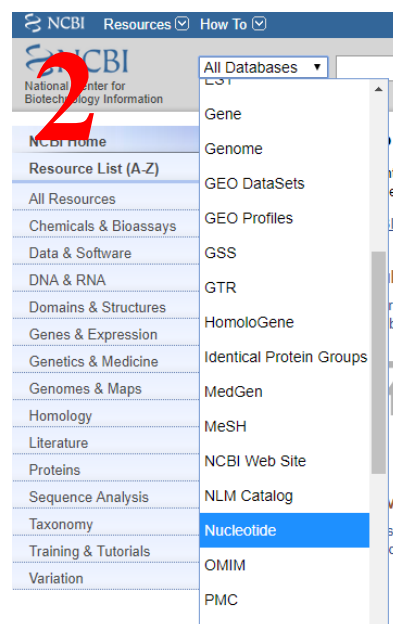
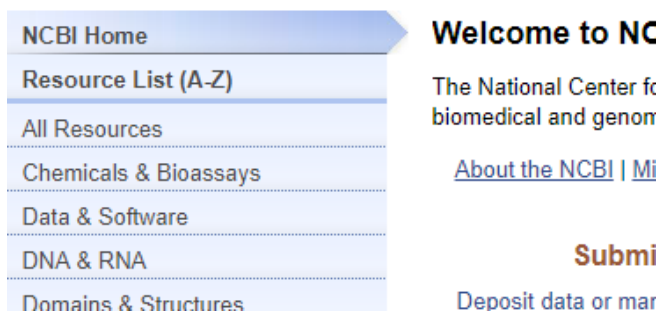
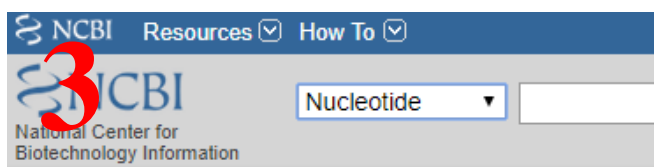
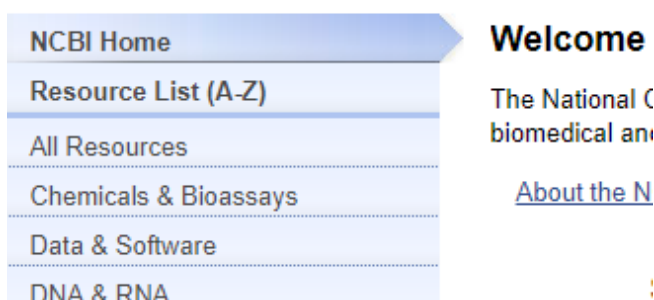
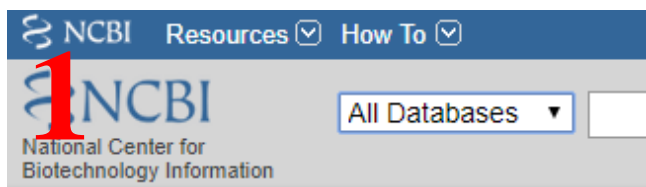




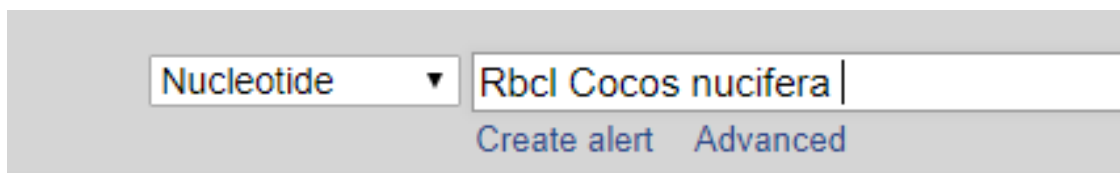
É válido mencionar que, como o NCBI é um programa norte-americano, os termos colocados na caixa de texto devem estar em inglês ou o nome da espécie pesquisada.

## Passo 2:

Altere a caixa seletora de banco para o banco de nucleotídeos, conforme figuras abaixo:



Em seguida coloque na caixa de texto o seguinte: ***Rbcl Cocos nucifera***



Provavelmente a página que aparecerá será a seguinte:







UNIVERSIDADE FEDERAL DE SERGIPE  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE - CCBS  
DEPARTAMENTO DE BIOLOGIA - DBI



NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Rbcl Cocos nucifera Search

Create alert Advanced Help

Species Plants (80) Customize ...

Molecule types genomic DNA/RNA (80) Customize ...

Source databases INSDC (GenBank) (79) RefSeq (1) Customize ...

Genetic compartments Chloroplast (77) Plastid (79)

Sequence length Custom range...

Release date Custom range...

Revision date Custom range...

Clear all

Show additional filters

Summary 20 per page Sort by Default order Send to:

See [rbcl, ribulose biphosphate carboxylase large chain](#) in the Gene database  
[rbcl](#) reference sequences [Protein \(1\)](#)

Items: 1 to 20 of 80

<< First < Prev Page 1 of 4 Next > Last >>

☐ [Cocos nucifera ribulose biphosphate carboxylase large subunit \(rbcl\) gene, complete cds; chloroplast gene for chloroplast product](#)  
1,445 bp linear DNA  
Accession: AY012507.1 GI: 13242846  
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

☐ [Cocos nucifera partial rbcl gene for ribulose-biphosphate carboxylase large subunit, specimen voucher 1968-4480 \(K\)](#)  
1,331 bp linear DNA  
Accession: HG969691.1 GI: 820653924  
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Cocos nucifera plastid partial rbcl gene for ribulose biphosphate carboxylase large subunit, specimen voucher 1968-4480 \(K\)](#)  
1,306 bp linear DNA  
Accession: AM110211.1 GI: 90968291  
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Cocos nucifera voucher 1764 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit](#)

Filters: [Manage Filters](#)

Find related data Database: Select Find items

Search details

Clique no primeiro link, onde você obterá algumas informações à cerca deste gene.

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Advanced Search Help

GenBank Send to:

**Cocos nucifera ribulose biphosphate carboxylase large subunit (rbcl) gene, complete cds; chloroplast gene for chloroplast product**

GenBank: AY012507.1

[FASTA](#) [Graphics](#) [PopSet](#)

Go to: ☺

LOCUS AY012507 1445 bp DNA linear PLN 14-JUL-2016

DEFINITION Cocos nucifera ribulose biphosphate carboxylase large subunit (rbcl) gene, complete cds; chloroplast gene for chloroplast product.

ACCESSION AY012507

VERSION AY012507.1

KEYWORDS .

SOURCE chloroplast Cocos nucifera (coconut palm)

ORGANISM [Cocos nucifera](#)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Areaceae; Arecoideae; Cocoseae; Attaleinae; Cocos.

REFERENCE 1 (bases 1 to 1445)

AUTHORS Hahn, W.J.

TITLE A Molecular Phylogenetic Study of the Palmae (Areaceae) Based on atpB, rbcl, and 18S nrDNA Sequences

JOURNAL Syst. Biol. 51 (1), 92-112 (2002)

PUBMED [11943094](#)

REFERENCE 2 (bases 1 to 1445)

AUTHORS Hahn, W.J.

TITLE Direct Submission

Change region shown

Customize view

Analyze this sequence Run BLAST Pick Primers

Nesta página você encontrará informações como Locus, tamanho do gene, se é gene linear ou circular, número de acesso, os autores que depositaram a sequência etc... Ao rolar a página para baixo você encontrará as letras CDS, que significa *coding DNA sequence* ou região codificante.

É comum encontrar na descrição do gene o termo "**partial cds**", este termo significa que a sequência do gene não está completa, pois quem depositou não obteve a sequências completa submetendo somente uma parte do gene. Já se a descrição for: "**complete cds**" que dizer que a sequência completa do gene foi depositada.







source 1..1445  
/organism="Cocos nucifera"  
/organelle="plastid:chloroplast"  
/mol\_type="genomic DNA"  
/specimen\_voucher="Hahn 7042, WIS, cult. FTG 80 798"  
/db\_xref="taxon:13894"

1 **gene** 2  
1..1434  
/gene="rbcL"

1 **CDS** 3  
1..1434  
/gene="rbcL"  
/codon\_start=1  
/transl\_table=11  
/product="ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit"  
/protein\_id="AAK14892.1"

4  
/translation="MSPQXETKASVGFAGVKDYKLTYYTPDXETKDTDILAAFRVTP  
QPGVPPEEAGAAVAEESTGTWTTVWTDGLTSLDRYKGRCYHIETVVGEENQYIAYVA  
YPLDLFEESVTNMFSTIVGNVFGFKALRALRLEDLRIPTSYSKTFQGPPIHQVERD  
KLNKYGRPLLGCITPKLGLSAKNYGRAVYECRLGGDLFTKDDENVNSQPFMRWRDRF  
CFCAEAIYKAQETGEIKGHYLNATAGTCEEMIKRAVFARELGAPIVMHDYLTGGFTA  
NTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSSGGDHIHAGTVVGK  
LGGEDNTLGEVNLDDDETEKDDPDECTEETDNLMDPDECTDVAECCTHAAHMDALTE

O que temos nesta imagem é:

- 1= CDS ou região codificante;
- 2= Gene cujo as informações serão descritas;
- 3= Produto do gene em questão;
- 4= sequência de aminoácidos (aa) do gene em questão.

Além destas informações existem várias outras que podem ser adquiridas, com por exemplo a sequência de nucleotídeos (nt) deste gene:

ORIGIN

```

1 atgtcaccac aarcagagac taaagcaagt gttggatttm aagctggtgt taaagattac
61 aaattgactt attatactcc tgactmcgaa accaaagata ctgatattctt ggcagcattc
121 cgagtaactc ctcaaccggg agttccgcct gaggaagcag gggcagcggg agctgcccga
181 tctttacttg gtacatggac aactgtgtgg actgatggac ttaccagtct tgatcggtac
241 aaaggacgat gctaccacat cgaaccggtt gtcggggagg aaaatcaata tattgcttat
301 gtagcctatc ctttagacct tttgaagaa ggttctgtta ctaacatggt tacttccatt
361 gtgggtaatg tatttggttt caaagcccta cgagctctac gtctggagga tctgcgaatt
421 cccacttctt attccaaaac tttccaaggc ccgctcatg gtatccaaat tgaagagat
481 aagttgaaca agtatggctg tctctattg ggatgtacta ttaaaccaaa attgggatta
541 tccgcaaaga actacggtag agcgggttat gaatgtctac gcggtggact tgattttacc
601 aaggatgatg aaaacgtgaa ctcacaacca tttatgcgtt ggagagaccg tttctgcttt
661 tgtgccgaag caatttataa agcgcaggcc gaaacgggtg aaatcaaagg acattacttg
721 aatgtacttg cgggtacatg tgaagaaatg atcaaaaggg ccgtatttgc cagagaattg
781 ggagctccta tcgtaatgca tgactactta actgggggat tcaactgaaa tactagcttg
841 gctcattatt gccgcgataa cggcctactt cttcacatcc atcgcgcaat gcatgcagtt
901 attgatagac agaaaaatca tggatgcat tttcgtgtac tagctaaagc attacgtatg
961 tctggtggag atcatattca cgcgggtaca gtagtgggta aactggaaagg ggaacgtgag
1021 atgacttttg gttttgttga tttattacgt gatgatttta ttgaaaaaga ccgaagtccg
1081 ggtatctttt ttactcaaga ttgggtctct atgccagggt ttataccgtt ggcttcaggg
1141 ggtattcatg tttggcatat gcctgccctg accgaaatct ttggggatga tttctgacta
1201 cagtttggcg gaggaacttt aggacacctt tggggaaatg caccgggtgc agtagcta
1261 cgggtggctt tagaagcgtg tgtacaagct cgtaatgaag gacgtgatct tgcccgtgaa
1321 ggtaaatgaaa ttatccgtga agctagcaaa tggagccctg aactagctgc cgcttgtaga
1381 gtatggaagg cgatcaaatt cgaattcgaa ccagtagata agctggataa atgaagagat
1441 aaaa

```





Todas estas informações são depositadas no banco de dados no formato Genbank, porém alguns programas usam estas sequências em outros formatos para executar uma série de análises, um destes formatos é o formato FASTA. Neste banco de dados você também consegue acessar esta sequência com o formato FASTA, basta ir no início da página e clicar em FASTA:

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Advanced Help

GenBank Send to

**Cocos nucifera ribulose biphosphate carboxylase large subunit (rbcL) gene, complete cds; chloroplast gene for chloroplast product**

GenBank: AY012507.1

[FASTA](#) [Graphics](#) [PopSet](#)

Go to: [v]

LOCUS AY012507 1445 bp DNA linear PLN 14-JUL-2016

DEFINITION Cocos nucifera ribulose biphosphate carboxylase large subunit (rbcL) gene, complete cds; chloroplast gene for chloroplast product.

ACCESSION AY012507

VERSION AY012507.1

KEYWORDS .

SOURCE chloroplast Cocos nucifera (coconut palm)

ORGANISM Cocos nucifera

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Areaceae; Arecoideae; Cocoseae; Attaleinae; Cocos.

REFERENCE 1 (bases 1 to 1445)

AUTHORS Hahn,W.J.

TITLE A Molecular Phylogenetic Study of the Palmae (Areaceae) Based on atpB, rbcL, and 18S nrDNA Sequences

JOURNAL Syst. Biol. 51 (1), 92-112 (2002)

PUBLISHED 11943094

REFERENCE 2 (bases 1 to 1445)

AUTHORS Hahn,W.J.

TITLE Direct Submission

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

**Cocos nucifera ribulose biphosphate carboxylase large subunit (rbcL) gene, complete cds; chloroplast gene for chloroplast product**

GenBank: AY012507.1

[FASTA](#) [Graphics](#) [PopSet](#)

Go to: [v]

Ao clicar em FASTA, você será encaminhado a uma página onde a sequência de nucleotídeos está em formato Fasta.

O formato FASTA inicia com o símbolo >, seguido do nº de acesso e descrição do gene.





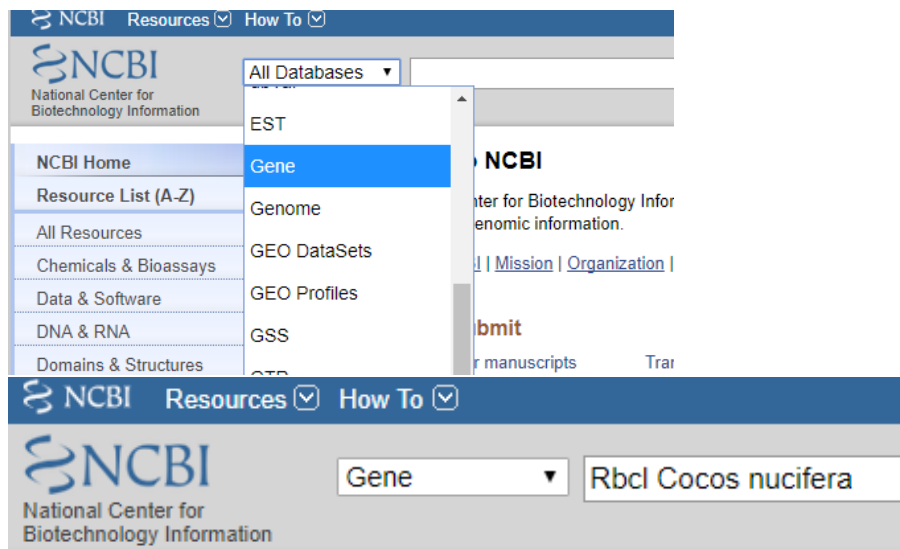
### Cocos nucifera ribulose biphosphate carboxylase large subunit (rbcL) gene, complete cds; chloroplast gene for chloroplast product

GenBank: AY012507.1

[GenBank](#) [Graphics](#) [PopSet](#)

```
>AY012507.1 Cocos nucifera ribulose biphosphate carboxylase large subunit (rbcL) gene,
complete cds; chloroplast gene for chloroplast product
ATGTCACCACAARCAAGACTAAAGCAAGTGTGGATTMAAGCTGGTAAAGATTACAAATTGACTT
ATTATACCTCTGACTMCGAAACCAAAGATACTGATATCTTGGCAGCATTCCGAGTAACCTCAACCCGG
AGTTCGCCCTGAGGAAGCAGGGGACGGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTGTGG
ACTGATGGACTTACAGCTCTTGATCGTTACAAAGGACGATGCTACACATCGAAACCGTTGTGGGGGAGG
AAATCAATATATTGCTTATGTAGCTATCCTTTAGACCTTTTGAAGAAGGTTCTGTTACTAATGTT
TACTTCCATTGTGGGTAATGTATTTGGTTTCAAAGCCTACGAGCTCTACGTCTGGAGGATCTGCGAATT
CCCATCTTCTATTCCAAACTTTCCAAAGGCCCGCTCATGGTATCCAAAGTTGAAAGAGATAAGTTGAACA
AGTATGGTCGTCTCTATTGGGATGTAATTAACCAAAATTTGGGATTATCCGCAAGAAGTACGGTAG
AGCGGTTTATGAATGTCTACGCGGTGGACTTGATTTTACCAAGGATGATGAAACGTGAACTCACACCA
TTTATGCGTTGGAGAGACCGTTTCTGCTTTTGTGCCGAAGCAATTTATAAGCGCAGGCCGAAACGGGTG
AAATCAAAGGACATTACTTGAATGCTACTGCGGGTACATGTGAAGAAATGATCAAAGGGCCGATTTCG
CAGAGAATTGGGAGCTCTATCGTAATGCATGACTACTTAAGTGGGGATTCACTGCAAACTAGCTTG
GCTCATTATTGCCGCGATAACGGCCTACTTCTTACATCCATCGCGCAATGCATGCAGTTATTGATAGAC
AGAAAAATCATGGTATGCATTTTCGTGTACTAGCTAAAGCATTACGTATGTCTGGTGGAGATCATATTCA
CGCGGGTACAGTAGTGGGTAACTGGAAAGGGGACGTGAGATGACTTTGGGTTTGTGATTATTACGT
GATGATTTTATTGAAAAAGACCGAAGTCGCGGTATCTTTTACTCAAGATTGGGTCTCTATGCCAGGTG
TTATACCGCTGGCTTCAGGGGTATTCTGTTTGGCATATGCCCTGCCCGAAATCTTTGGGGATGA
TTCTGTACTACAGTTTGGCGGAGGAACCTTAGGACACCTTGGGGAAATGCACCGGTGCAGTAGTAAT
CGGGTGGCTTTAGAAAGCTGTGTACAAGCTCGTAATGAAGGACGTGATCTTGCCCGTGAAGGTAATGAAA
TTATCCGTGAAGCTAGCAAAATGGAGCCCTGAAGTGTGCGCTTGTGAAGTATGGAAGGCGATCAAAAT
CGAATTCGAACCAAGTAGATAAGCTGGATAAATGAAGAGATAAAAT
```

Outra maneira de adquirir informações sobre genes é acessando o banco de dados de gene alterando a caixa seletora de banco de dados, e pesquisando pelo gene de interesse, seja na busca convencional ou na busca avançada.



Na busca convencional você será redirecionado à uma página com informações do gene pesquisado:





NCBI Resources How To

Gene Gene Rbcl Cocos nucifera  
Create RSS Create alert Advanced

Full Report Send to

Showing Current items.

**rbcl ribulose biphosphate carboxylase large chain [ *Cocos nucifera* (coconut palm) ]**  
Gene ID: 17046384, updated on 5-Nov-2016

**Summary**

Gene symbol: rbcl  
Gene description: ribulose biphosphate carboxylase large chain  
Locus tag: Q438\_p061  
Gene type: protein coding  
RefSeq status: PROVISIONAL  
Organism: [Cocos nucifera \(cultivar: dwarf coconut\)](#)  
Lineage: Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Areaceae; Arecoideae; Cocoseae; Attaleinae; Cocos

**Genomic context**

Sequence: NC\_022417.1 (54542..55996)

Diagram showing the genomic context of the rbcl gene on the NC\_022417.1 sequence. The gene is located between the atpB and accD genes. The diagram shows the following features: trnH-GRU, atpB, rbcl, accD, and psal. The rbcl gene is highlighted in red.

### Passo 3

Para realizar uma busca avançada, vá ao banco de dados de gene como anteriormente e clique em “Advanced”.

NCBI Resources How To

Gene Gene

Advanced

Gene

Gene integrates informa

Você será encaminhado a uma outra página com os seguintes campos:

### Builder

All Fields

AND All Fields

Search or [Add to history](#)





Em “**All Fields**” você delimita em que campo você deseja pesquisar, por exemplo, ao clicar no campo **Gene name**, você diz ao software que gostaria de pesquisar pelo nome do gene, ao clicar em “**AND**” você diz ao software que operador você deseja ativar nesta busca, sendo eles:

**AND** – retorna ao usuário apenas resultados que obtenham os termos colocados em pesquisa.

Exemplo:

(Rbcl[Gene Name]) AND cocos nucifera[Organism]

[Edit](#)

#### Builder

Gene Name	▼	Rbcl	
AND ▼	Organism	▼	cocos nucifera

[Search](#) or [Add to history](#)

Neste caso serão retornados resultados que correspondam ao gene de nome “Rbcl” E ao organismo “*Cocos nucifera*”, logo resultados de Rbcl para outros organismos como *Atleae funifera* por exemplo, não irão aparecer.

**OR**- retorna ao usuário apenas resultados que obtenham os termos de um campo ou outro.

Exemplo:

(Rbcl[Gene Name]) OR Cocos nucifera[Organism]

[Edit](#)

#### Builder

Gene Name	▼	Rbcl	
OR ▼	Organism	▼	Cocos nucifera

[Search](#) or [Add to history](#)

Neste caso tanto resultados correspondentes à “Rbcl” quanto à “*Cocos nucifera*” terão seus resultados retornados.





How To Sign in to NCBI

Gene

Create RSS Create alert Advanced Help

Tabular 20 per page Sort by Relevance Send to: Hide sidebar >>

**Search results**

Items: 1 to 20 of 273

See also 152 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases
<a href="#">Rbcl</a> ID: 17046384	ribulose biphosphate carboxylase large chain [Cocos nucifera (coconut palm)]	NC_022417.1 (54542..55996)	Q438_p061
<a href="#">Rbcl</a> ID: 2717040	RuBisCO large subunit [ <i>Chlamydomonas reinhardtii</i> ]	NC_005353.1 (122490..123917, complement)	ChreCp049
<a href="#">Rbcl</a> ID: 844754	RuBisCO large subunit [ <i>Arabidopsis thaliana</i> (thale cress)]	NC_000932.1 (54958..56397)	ArthCp030
<a href="#">Rbcl</a> ID: 845212	RuBisCO large subunit [ <i>Zea mays</i> ]	NC_001666.2 (56874..58304)	ZemaCp032
<a href="#">Rbcl</a> ID: 800513	RuBisCO large subunit [ <i>Nicotiana tabacum</i> (common tobacco)]	NC_001879.2 (57600..59033)	NitaCp031
<a href="#">Rbcl</a> ID: 4126887	RuBisCO large subunit [ <i>Oryza sativa Indica Group</i> (long-grained rice)]	NC_008155.1 (54030..55484)	OrsaCp023
<a href="#">Rbcl</a> ID: 3131463	RuBisCO large subunit [ <i>Oryza sativa Japonica Group</i> (Japanese rice)]	NC_001320.1 (54095..55528)	OrsaCp033
<a href="#">Rbcl</a> ID: 4099985	RuBisCO large subunit [ <i>Solanum tuberosum</i> (potato)]	NC_008096.2 (56531..57964)	SotuCp029

clear

**Results by taxon**

Top Organisms [\[Tree\]](#)

- Cocos nucifera (263)
- Oryza sativa (4)
- Olea europaea (4)
- Saccharum hybrid cultivar (3)
- Triticum aestivum (2)
- All other taxa (2457)

More...

**Find related data**

Database:

**Search details**

Rbcl[Gene Name] OR "Cocos nucifera"  
[Organism] AND alive[prop]

NOT - retorna ao usuário apenas resultados que obtenham os termos de um campo excluindo os resultados do termo que está no segundo campo.

Exemplo:

[Edit](#)

**Builder**

or [Add to history](#)

Neste caso todos os resultados correspondentes à Rbcl que não incluíam “*Cocos nucifera*” retornaram ao usuário.





How To

Gene (Rbcl[Gene Name]) NOT Cocos nucifera[Organism]

Create RSS Create alert Advanced

Tabular 20 per page Sort by Relevance Send to:

**Search results**

Items: 1 to 20 of 2470 << First < Prev Page 1 of 124 Next > Last >>

See also [152 discontinued or replaced items](#)

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> <a href="#">rbcl</a> ID: 2717040	RuBisCO large subunit [ <i>Chlamydomonas reinhardtii</i> ]	NC_005353.1 (122490..123917, complement)	ChreCp049
<input type="checkbox"/> <a href="#">rbcl</a> ID: 844754	RuBisCO large subunit [ <i>Arabidopsis thaliana</i> (thale cress)]	NC_000932.1 (54958..56397)	ArthCp030
<input type="checkbox"/> <a href="#">rbcl</a> ID: 845212	RuBisCO large subunit [ <i>Zea mays</i> ]	NC_001666.2 (56874..58304)	ZemaCp032
<input type="checkbox"/> <a href="#">rbcl</a> ID: 800513	RuBisCO large subunit [ <i>Nicotiana tabacum</i> (common tobacco)]	NC_001879.2 (57600..59033)	NitaCp031
<input type="checkbox"/> <a href="#">rbcl</a> ID: 4126887	RuBisCO large subunit [ <i>Oryza sativa Indica Group</i> (long- grained rice)]	NC_008155.1 (54030..55484)	OrsaiCp23
<input type="checkbox"/> <a href="#">rbcl</a>	RuBisCO large subunit [ <i>Oryza sativa Japonica Group</i> ]	NC_001320.1	OrsajCp033

## Passo 4

Acesse as informações desta proteína usando o número de acesso a seguir através do banco de dados de gene: **KX028884.1**

Agora responda as seguintes questões:

- 1) Qual foi o gene localizado através deste número de identidade?
- 2) A que organismo pertence este gene? E qual o filo, família e subfamília deste organismo?
- 3) O gene localizado codifica alguma proteína?
- 4) Qual título do artigo relacionado à deposição desta sequência?
- 5) colete a sequência de nt deste gene no formato FASTA?

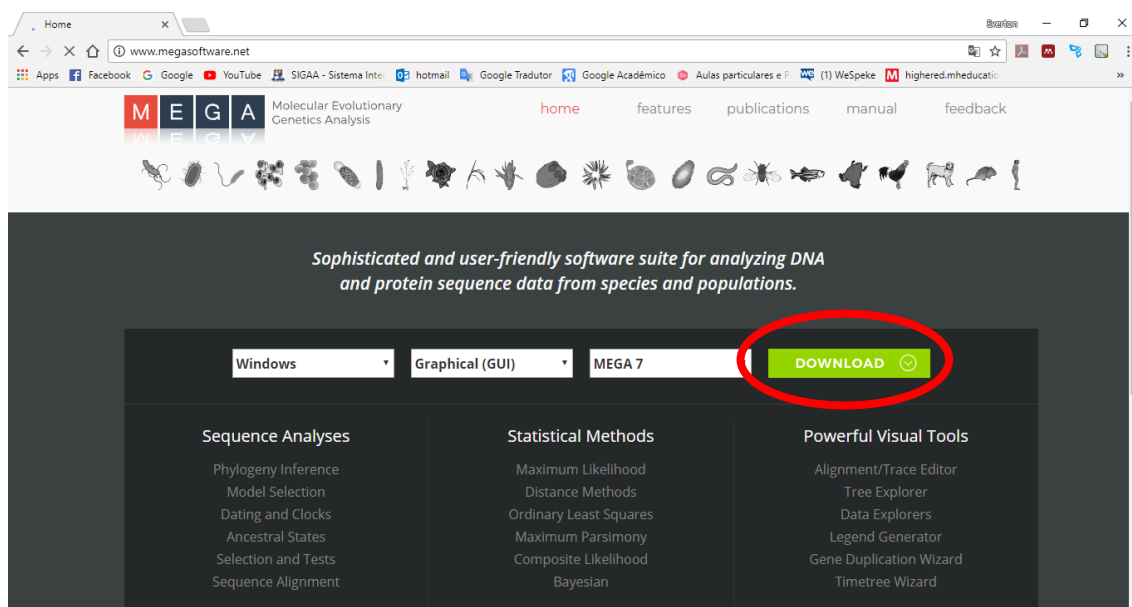




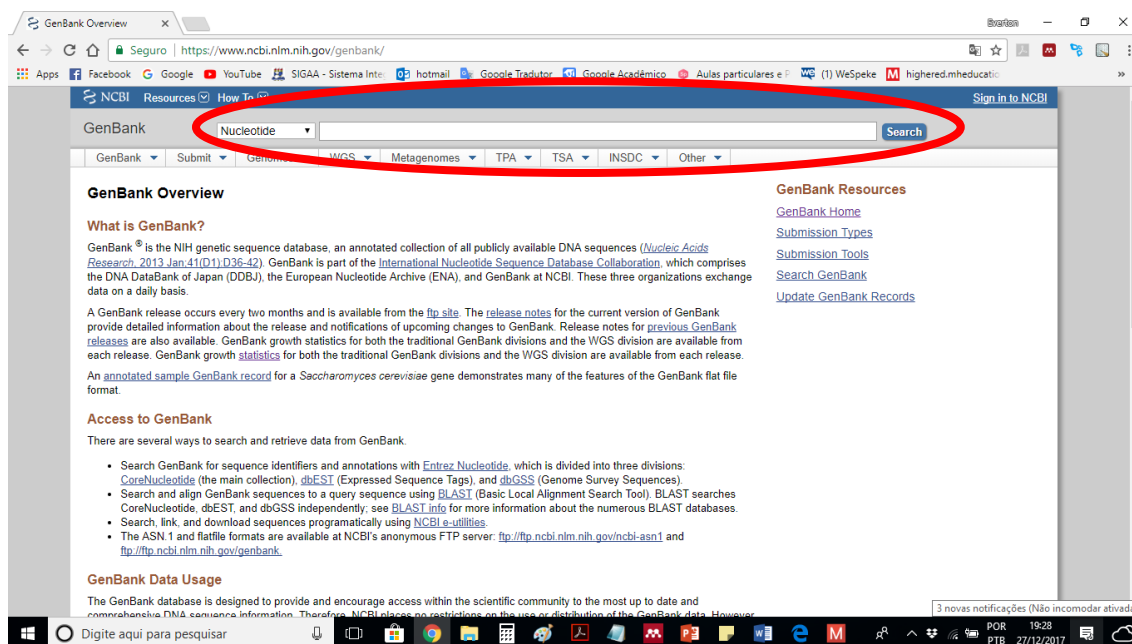


## TUTORIAL ALINHAMENTO

Para realizar este alinhamento iremos utilizar o programa MEGA 7, este encontra-se disponível para Download no site: < <http://www.megasoftware.net/> >.



Para realizar este alinhamento iremos adquirir as sequências no banco de dados GenBank, você pode acessar este banco de dados através do site: < <https://www.ncbi.nlm.nih.gov/genbank/> >.







UNIVERSIDADE FEDERAL DE SERGIPE  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE - CCBS  
DEPARTAMENTO DE BIOLOGIA - DBI



No campo de busca coloque o termo: “Human papillomavirus type 33 E6 complete Genome”, algo semelhante a imagem abaixo deve aparecer.

Human Papillomavirus Type 33 E6 Complete genome

Search

Species: Viruses (27)

Molecule types: genomic DNA/RNA (27)

Source databases: INSDC (GenBank) (27)

Sequence length: Custom range...

Release date: Custom range...

Revision date: Custom range...

Clear all

Show additional filters

Summary: 20 per page Sort by Default order

Send to: Filters: Manage Filters

Find related data: Database: Select

Search details: ("Human papillomavirus type 33" [Organism] OR Human Papillomavirus Type 33[All Fields]) AND ("Echovirus E6" [Organism] OR E6[All Fields]) AND Complete[All Fields] AND genome[All Fields]

Recent activity: Turn Off Clear

Human Papillomavirus Type 33 E6 Complete genome (27)

Phylogenetic classification of human papillomavirus genotypes in high-grade

Taxonomy Links for Nucleotide (Selected) Rio de Janeiro Com períodos nublados 24-31°C

602219922 (1)

Human papillomavirus type 33 isolate LZcc12-33, complete genome

7,833 bp circular DNA

Accession: KU298892.1 GI: 218931422

Protein

GenBank FASTA Graphics PopSet

Human papillomavirus type 33 isolate 65B.33, complete genome

7,909 bp circular DNA

Accession: KU298892.1 GI: 966752192

Protein

GenBank FASTA Graphics

Human papillomavirus type 33 isolate 65A.33, complete genome

7,909 bp circular DNA

Accession: KU298891.1 GI: 966752183

Protein

GenBank FASTA Graphics

Pode clicar no primeiro link, neste momento você está obtendo informações acerca do HPV tipo 33 e para fim desta atividade iremos usar o gene E6.

Para acessar as informações apenas deste gene, é necessário selecionar a região codificante deste, isto pode ser feito ao localizar a descrição do gene e clicando no termo “cds”, em seguida em “FASTA” para obter os dados em arquivo com o formato Fasta. Veja as imagens abaixo:

Human Papillomavirus Type 33 E6 Complete genome

Search

Species: Viruses (27)

Molecule types: genomic DNA/RNA (27)

Source databases: INSDC (GenBank) (27)

Sequence length: Custom range...

Release date: Custom range...

Revision date: Custom range...

Clear all

Show additional filters

Summary: 20 per page Sort by Default order

Send to: Filters: Manage Filters

Find related data: Database: Select

Search details: ("Human papillomavirus type 33" [Organism] OR Human Papillomavirus Type 33[All Fields]) AND ("Echovirus E6" [Organism] OR E6[All Fields]) AND Complete[All Fields] AND genome[All Fields]

Recent activity: Turn Off Clear

Human Papillomavirus Type 33 E6 Complete genome (27)

Phylogenetic classification of human papillomavirus genotypes in high-grade

Taxonomy Links for Nucleotide (Selected) Rio de Janeiro Com períodos nublados 24-31°C

602219922 (1)

Human papillomavirus type 33 isolate LZcc12-33, complete genome

7,833 bp circular DNA

Accession: KU298892.1 GI: 218931422

Protein

GenBank FASTA Graphics PopSet

Human papillomavirus type 33 isolate 65B.33, complete genome

7,909 bp circular DNA

Accession: KU298892.1 GI: 966752192

Protein

GenBank FASTA Graphics

Human papillomavirus type 33 isolate 65A.33, complete genome

7,909 bp circular DNA

Accession: KU298891.1 GI: 966752183

Protein

GenBank FASTA Graphics





UNIVERSIDADE FEDERAL DE SERGIPE  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE - CCBS  
DEPARTAMENTO DE BIOLOGIA - DBI



CVGLEIGRQPLGVISGHPLLNKFDDTETSNKYPGQGDNRCLSDYKQTLCLL  
GCKPPTGEHUGKGVACTNAAPANDCPPLLEINTIIEGDIVDTGFCDFKTLQANKS  
DVPIDICGSKYDYLKHTSEPYGDSLFFFLRREQIFVRHFFNRAGLGEAVPDLY  
IKGSGTTASIQSSAFFPTPSGSIWSESQLFNKPVYLQRAQHNGICGNQVPTVV  
DITRSTWITLCLTQVSDSTYKHNENKEVIRHVEYDLQVPLQCKVLTADVHTYIHA  
HNPISLEDDQGLTPPPSASLQDTRVPTVSQATCTQKTVPPKEEDPLQKVTTFUEVDL  
KEKFSADLDQPLGRKFLQAGLKAKPKLRAAPTSTRSSAKKKVKK"

ORIGIN

```
1 gtaaacata atgccaaagt ttaaaaaagt aggggtgaac cgaagcggt tcaaccgaa
61 acggtgcata tataaagcaa ccttttga gtaaggctc gcacgacat gtttcaagac
121 actgagpaaa aacacagac attgctgat ttgtccaaag cattgagac aactatcac
181 aacattgaac taagtgcat ggaatgaga aacctttgc aagatctga gataatgat
241 ttgtcctttg cagatttaac agttgtatat agggagagaa atccatttgg aatagttaa
301 ttgttcttgc gtttctatc taaaattagt gaatatagac attataatta ttctgtatat
361 agacatacat tagaacaac agttaacaa ctttaaatg aatatataat taggtgtatt
421 ttatgtcaaa gacctttgtg tcttcagaa aaaaaagc atgtggattt aaacaarga
481 ttatcagata ttctgatat aatgagaaa atcagagaa atcagagaa atcagagaa
541 ttatcagata ttctgatat aatgagaaa atcagagaa atcagagaa atcagagaa
601 aatatgtttt agattatat cctgaaccaa ctgacctata ctgtatgag caattaagt
661 acagctcaga tggagatgaa ggccttgacc gccagatgg acaagcaca ccagccacg
721 ctgtactata cattgaacc ttgtgtaca ctgtaacac cacagttcgt ttatgttca
781 acagtacagc aagtgaacta cgaacacat agcaactat tatgggcaca tgaatatgt
841 ttgtgctcac ctgtgcaca ctaaacatc cacttcaat ggcgtacct gaagtlcac
901 atggagctgg gtaggggtgt actgtgtgc ttgagtaga agcagctata gagagagaa
961 caggagataa tattcagaa gatgagatg aacagcaga tgacagtgc acggttacc
1021 tagagtttat agatgtctt atgaaaata gtatacagg agacacagag cagcccgga
1081 cattgtttta tacaagaaa gggagagatg attaaatgc ttgtgttga ctaaaagaa
1141 agtttgcgc atgttcacaa agtgcgagg aggaactgtg tgatgtgct gcaaacccgt
1201 gtgagagctc tattataaa aataaagat gcatcacag aaacagaaa atagtgagc
1261 tagaagacag cagatagtc aatcagag tggaaacta cgaagtgta caacagtag
1321 aaagtcaaaa tggcagaca aactaaatg acttagaatc tagtggggtg ggggatgatt
1381 cagaagtaag ctgtgagaca aatgtagata gctgtgaaa ttgtacttg caggaatta
1441 gtaagtgttc acatagtagt aatacaaaa caatatatt atataaatt aaagagcct
1501 atggagtaag ttttatggaa ttagtaagac catttaaaag tgataaaca agctgtacag
1561 atgggtgatg aacagagatg ggaattagtc catagtagc agaaagttta aagttatga
1621 ttaaacagca tagtttgtat actcatttac aatgttaac gtcgtaga ggaataata
1681 tattattgtt aattagattt aggtgtaga aaaaagctt aacagtagca aactaata
```

A região destacada é a região codificante do gene em questão.

109..558  
/gene="E6"  
/codon\_start=1  
/product="E6"  
/protein\_id="E6"  
/translation="MFQTEEL...  
FAFADLVVREGNPFICKLRL...  
CIIQRLCPREKKRVDLNRHITS...  
...SRRETAL"

CDS Feature 1 of 8 EU918766: 1 segment

Details FASTA GenBank Help

Human papillomavirus type 33 isolate LZcc12-33, complete genome

GenBank: EU918766.1

FASTA

Send to: Change region shown

Whole sequence  
Selected region  
from: 109 to: 558  
Update View

Customize view

Analyze this sequence  
Run BLAST  
Pick Primers

Genomic Neighbours

Aqui você tem apenas a região codificante do gene E6 no formato FASTA, ou seja, as informações das outras regiões não influenciarão na sua análise.

Para obter Regiões similares basta clicar em “Run BLAST” no lado esquerdo da tela que você será encaminhado à uma página para a realização do BLAST.





UNIVERSIDADE FEDERAL DE SERGIPE  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE - CCBS  
DEPARTAMENTO DE BIOLOGIA - DBI



Human papillomavirus type 33 isolate LZcc12-33, complete genome

GenBank: EU918766.1

FASTA

Human papillomavirus type 33 isolate LZcc12-33, complete genome

GenBank: EU918766.1

ATGTTTCAAGACACTGAGGAAAAACACGACATTGTCATGATTGTGCGCAAGCATTGGAGCAACTATAC  
ACACATTGACACTACAGTGCSTGGATGCAAGAACTTTGCAAGCATCTGAAGTATATGATTGTCATT  
TGCAGATTAAACAGTTGTATATAGGAGGGAATCCATTGGGAATATGTAACGTGTGTTGCGGTTCTTA  
TCTAAATTAAGTGAATATAGACATTATAATTCTGTATATGGACATACATAGAACAAACAGTTAAACA  
AACCTTTAAATGAAATATTAATTAGGTGATATATATGTCAGAACCTTTGTGCTCGAGAAAAAAGCG  
ACATGTGGATTAAACAAACGATTTCATAATATTCGGGTCGTTGGGCAAGGCGCTGTGCGGCGTGTGG  
AGGTCCGACGTAGAGAACTGCACGTGA

Change region shown

Whole sequence  
Selected region

from: 109 to: 558

Update View

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Find in this Sequence

Related information

Protein

BLAST

Standard Nucleotide BLAST

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

EU918766.1

Query subrange

From: 109 To: 558

Or, upload file

Escolher arquivo Nenhum arquivo selecionado

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Human genomic + transcript Mouse genomic + transcript Others (nr etc.)

Nucleotide collection (nr/nt)

Organism

Optional

Enter organism name or id-completions will be suggested

Exclude

Optional

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude

Optional

Models (XM/XP) Uncultured/environmental sample sequences

Limit to

Optional

Sequences from type material

Entrez Query

Optional

Enter an Entrez query to limit search

Agora, alguns parâmetros devem ser modificados a fim de permitir a localização de sequências menos similares (porém ainda homólogas) às suas.

Para tal, role a página para baixo e clique em: “Somewhat similar sequences (blastn)”, ao clicar neste parâmetro, você está permitindo que o programa busque por sequências menos similares à sua.

Em seguida clique em “Algorithm parameters” para expandir alguns parâmetros que ficam ocultos, ao fazer isto, selecione o parâmetro “Max target sequences” e o modifique para 1000, assim você permite que o programa lhe retorne 1000 alvos encontrados ao invés de apenas 100.

Veja as imagens abaixo:





UNIVERSIDADE FEDERAL DE SERGIPE  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE - CCBS  
DEPARTAMENTO DE BIOLOGIA - DBI



Optional

Enter organism name or id—completions will be suggested

Exclude

Optional

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Models (XM/XP) ☐ Uncultured/environmental sample sequences

Sequences from type material

Entrez Query

Optional

Enter an Entrez query to limit search

Program Selection

Choose a BLAST algorithm

☐ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☒ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST

Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

Algorithm parameters

General Parameters

Max target sequences: 100

Select the maximum number of aligned sequences to display

Short queries: ☒ Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: 11

Max matches in a query range: 0

Nucleotide BLAST: Search

Seguro

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides&PROGRAM=blastn&QUERY=EU918766.1&DATABASE=nr&MEGABLAST=on&BLAST\_PROG...

More dissimilar sequences (discontiguous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST

Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign

Restore default search parameters

General Parameters

Max target sequences: 1000

Select the maximum number of aligned sequences to display

Short queries: ☒ Automatically adjust parameters for short input sequences

Expect threshold: 100

Word size: 250

Max matches in a query range: 1000

Scoring Parameters

Match/Mismatch Scores: 5000

Gap Costs: 5 Extension: 2

Filters and Masking

Filter: ☒ Low complexity regions

☐ Species-specific repeats for: Homo sapiens (Human)

Mask: ☒ Mask for lookup table only

☐ Mask low complexity letters

Feito isto, clique em BLAST e aguarde os resultados.

Max matches in a query range: 0

Scoring Parameters

Match/Mismatch Scores: 2, -3

Gap Costs: Existence: 5 Extension: 2

Filters and Masking

Filter: ☒ Low complexity regions

☐ Species-specific repeats for: Homo sapiens (Human)

Mask: ☒ Mask for lookup table only

☐ Mask low complexity letters

BLAST

Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)

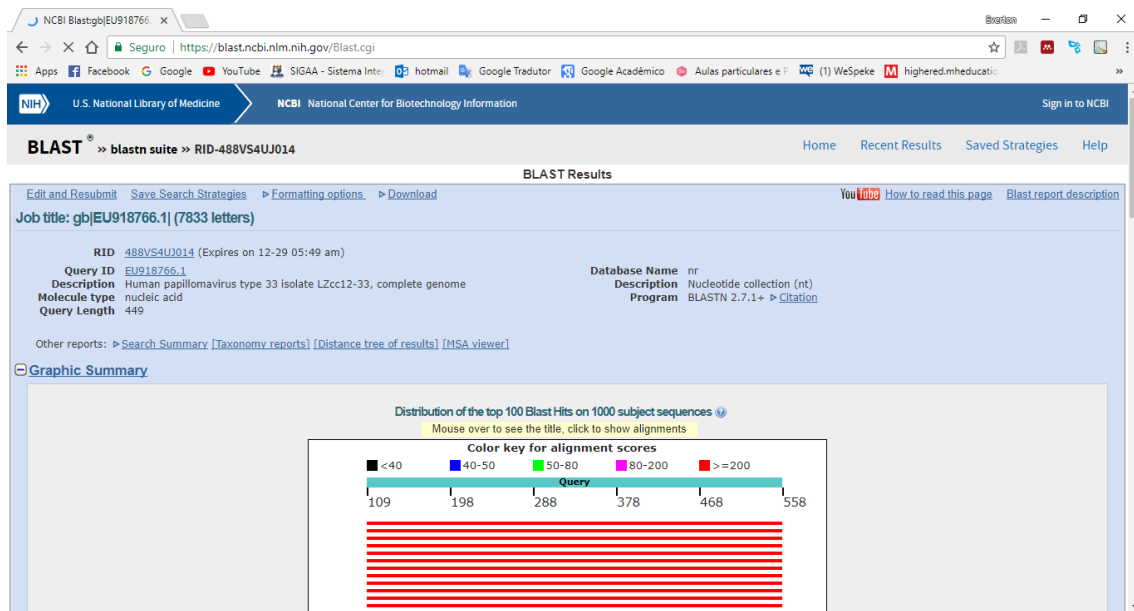
Show results in a new window

Ao finalizar o programa lhe retornará uma página semelhante a esta:





UNIVERSIDADE FEDERAL DE SERGIPE  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE - CCBS  
DEPARTAMENTO DE BIOLOGIA - DBI



Nela, você consegue observar algumas informações sobre os hits encontrados.

Role a página para baixo e selecione 5 tipos de papilomavírus diferentes, para facilitar use os números de acesso seguintes: [EU918766.1](#) , [KT799935.1](#), [KY225962.1](#), [KU298930.1](#), [KF700140.1](#).

Estes hits escolhidos apresentam o genoma completo ou mais de um gene do tipo de papilomavírus, logo, deve clicar no seu número de acesso para obter informações apenas do gene E6.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Human papillomavirus type 33 isolate L2cc12-33, complete genome	812	812	100%	0.0	100%	<a href="#">EU918766.1</a>
Human papillomavirus type 33 isolate 33HE-15 E6 and E7 genes, complete cds	807	807	100%	0.0	99%	<a href="#">KC354758.1</a>
Human papillomavirus type 33 pop-variant IARC variant 15 early protein E6 (E6) and early protein E7 (E7) genes, complete cds	807	807	100%	0.0	99%	<a href="#">KC862078.1</a>
Human papillomavirus type 33 isolate 33HE-14 E6 and E7 genes, complete cds	794	794	100%	0.0	99%	<a href="#">KC354757.1</a>
Human papillomavirus type 33 isolate 33HE-13 E6 and E7 genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC354756.1</a>
Human papillomavirus type 33 isolate 33HE-12 E6 and E7 genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC354755.1</a>
Human papillomavirus type 33 isolate 149 E6 protein (E6) gene, complete cds	789	789	100%	0.0	99%	<a href="#">KF700161.1</a>
Human papillomavirus type 33 pop-variant IARC variant 12 early protein E6 (E6) and early protein E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC881016.1</a>
Human papillomavirus type 33 pop-variant IARC variant 10 early protein E6 (E6) and early protein E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC881015.1</a>
Human papillomavirus type 33 pop-variant IARC variant 9 early protein E6 (E6) and early protein E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC881014.1</a>
Human papillomavirus type 33 pop-variant IARC variant 8 early protein E6 (E6) and early protein E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC881013.1</a>
Human papillomavirus type 33 pop-variant IARC variant 7 early protein E6 (E6) and early protein E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC881012.1</a>
Human papillomavirus type 33 pop-variant IARC variant 14 early protein E6 (E6) and early protein E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC862077.1</a>
Human papillomavirus type 33 pop-variant IARC variant 13 early protein E6 (E6) and early protein E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">KC862076.1</a>
Human papillomavirus type 33 isolate B597 E6 (E6) gene, complete cds	789	789	100%	0.0	99%	<a href="#">KC662565.1</a>
Human papillomavirus type 33 isolate S80981 E6 (E6) and E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">JQ976781.1</a>
Human papillomavirus type 33 isolate S42451 E6 (E6) and E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">JQ976770.1</a>
Human papillomavirus type 33 isolate S42091 E6 (E6) and E7 (E7) genes, complete cds	789	789	100%	0.0	99%	<a href="#">JQ976767.1</a>
Human papillomavirus type 33 isolate Qv32431, complete genome	789	789	100%	0.0	99%	<a href="#">HQ537701.1</a>
Human papillomavirus type 33 isolate IN109456, complete genome	789	789	100%	0.0	99%	<a href="#">HQ537700.1</a>

Ao clicar no número de acesso, clique novamente na região CDS do gene E6 e clique em FASTA para obter as informações desta região, ao fazer isto, selecione as informações, copie e cole-as num bloco de notas **dando “um enter”** de uma sequência para outra.





UNIVERSIDADE FEDERAL DE SERGIPE  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE - CCBS  
DEPARTAMENTO DE BIOLOGIA - DBI



NCBI Blastgb/EU918766.1 x Human papillomavirus t: x

Seguro | https://www.ncbi.nlm.nih.gov/nuccore/EU918766.1?from=109&to=558&sat=4&sat\_key=66519832&report=fasta

NCBI Resources | How To | Sign in to NCBI

Nucleotide Nucleotide Search Help

FASTA

Human papillomavirus type 33 isolate LZcc12-33, complete genome

GenBank: EU918766.1

GenBank Graphics PopSet

Human papillomavirus type 33 isolate LZcc12-33, complete genome

ATGTTTCAAGACACTGAGGAAAAACACGACATTCATGATTTGTGCAAGCATTGGAGACACTATAC  
ACAACATTGAACACAGTGTGCGTGAATGCAGAAACCTTTGCAACGATCTGAGGTATATGATTTGCAAT  
TCGAGATTAAACAGTTGATATAGGGAGGGAAATCCATTGGAAATGTAACATGTTTGCAGTTCTTA  
TCTAAATTAGTGAATATAGACATTATAATTCTGTATATGACATACATTAGAACAAACAGTTAAACA  
ATCTTTAAATGAAATATATATAGGTGATTTAGGTGATTTATATGTCAGAACCTTTGTCTCTCGAGAAAAACG  
ACATGTGGATTAAACAAACGATTTCAATATATTTTCGGGTCTTTGGGACGGGCTGTGCGGCTGTTGG  
AGGTCTCCGACGTAAAGAAATGACACTGTGA

Change region shown

Whole sequence  
Selected region  
from: 109 to: 558  
Update View

Customize view

Analyze this sequence

Run BLAST  
Pick Primers  
Find in this Sequence

Related information

in  
ed  
ed  
onomy  
Full text in PMC  
Genomic Neighbours

Copie a sequência e o título dela e salve num bloco de notas.

filonegia HPV\_teste Everton.txt - Bloco de notas

Arquivo Editar Formatar Exibir Ajuda

152739.1:22-477 Alphapapillomavirus 9 isolate AsAme5 E6 protein gene, complete cds

ATGTTTCCAGGACCCAGGAGGACCGACCCAGAAAGTTACACATTATGACACAGAGCTGCAAAACACTATAC  
ATGATATAATATTAGCAATGTGTACTGCAAGCAACAGTTACTGCGACGTGAGGTATATGACATTTGCTTT  
TCGGGATTTATGCATAGTATATAGAGATGGGAATCCATATGCAAGTGTGATAAATGTTAAAGTTTAT  
TCTAAATTAGTGAGTATAGATATATTGTTATAGTGTGTATGGAACAACTTAGAACAGCAATACAACA  
AACCGTTGTGTGATTTGTTAATAGGTGATTAACCTGCAAAAGCCACTGTGCTCGAAGAAAAAGCAAG  
ACATCTGGACAAAAAGCAAGATTCCATAATATAAGGGCTCGGTGGACCGGTGATGATGCTTTGTTGG  
AGATCATCGAGAACCTGAGAGAAACCCAGCTGTAA

>KT799935.1:1-447 Human papillomavirus type 52 isolate 52HE26 early protein (E6) and early protein (E7) genes, complete cds

ATGTTTGAAGGATCCAGCAACACGACCCCGGACCCCTGCACGAATTTGTGTGAGGTGCTGGAAGAATCGGTGC  
ATGAAATAGGCTGCAAGTGTGCAAGTGCAGAAAGAGCTACAACGAAGAGAGGTATACAAGTTTCTATT  
TACAGATTACGAATAGTATATAGAGACAATATCCATATGGCGGTGTGATTTATGTCCTACGCTTTTATA  
TCTAAGATAAGTGAATATAGGCATTATCAATATTCACTTTATGGAAGAAACATTAGAAGAAAGGGTAAGAA  
AACCAATTAAGTGAATTAATATAGATGTAATTTGTCAACGCCAATTATGTCCTGAAGAAAAAGAAAG  
ACATGTTAATGCAAAACAGCGATTTCAATATATTGGGTCTGTGGACAGGGCGCTGTCAGAGTGTGG  
AGACCCGACCTGTGACCAAGTGTAA

>EU918766.1:109-558 Human papillomavirus type 33 isolate LZcc12-33, complete genome

ATGTTTCAAGACACTGAGGAAAAACACGACATTCATGATTTGTGCAAGCATTGGAGACACTATAC  
ACAACATTGAACACAGTGTGCGTGAATGCAGAAACCTTTGCAACGATCTGAGGTATATGATTTGCAAT  
TCGAGATTAAACAGTTGATATAGGGAGGGAAATCCATTGGAAATGTAACATGTTTGCAGTTCTTA  
TCTAAATTAGTGAATATAGACATTATAATTCTGTATATGACATACATTAGAACAACAGTTAAACA  
AACCTTTAAATGAATATTAATAGGTGATTTATATGTCAGAACCTTTGTGCTCGAGAAAAACG  
ACATGTGGATTAAACAACGATTTCAATATATTGGGTCTGTGGACAGGGCGCTGTGCGGCTGTTGG  
AGGTCCGACGTAGAGAACTGCACGTGA

>KY225962.1:110-559 Human papillomavirus type 58 isolate ZWE064490, complete genome

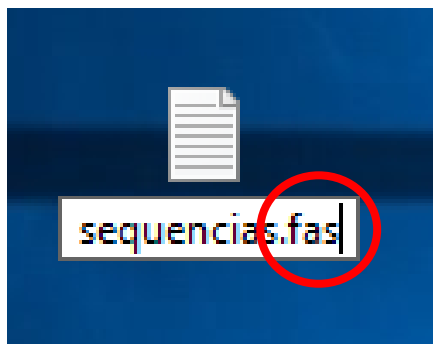
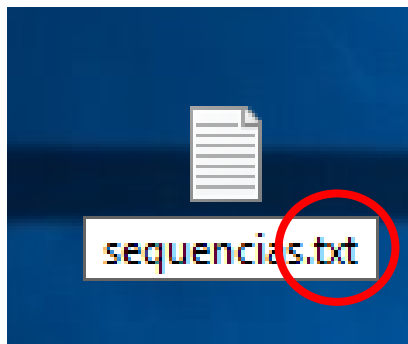
ATGTTTCCAGGACCGAGAGGAGAAACACGACATTCATGATTTGTGTCAGGCGTTGGAGACATCTGTGC  
ATGAAATGGAATGAATGCGTTGAATGCAAAAGAGCTTTGACGGATCTGAGGTATATGACTTTGTATT  
TCGACATTAAAGAAATAGTGTATAGAGATGGAAATCCATTGCAAGTATGAAAGTGTGTTACGATTGCTA  
TCTAAATAGTGAGTATAGACATTATAATTATCGCTATATGGAGACACATTAGAACAACACTAAAAA  
AGTGTTTAATGAATATTAATAGATGTAATTTGTCAAGACCAATTGTGTCACAAAGAAAAAAG  
GCATGTGGATTAAACAAGGTTTCAATATATTGGGTCTGTGGACAGGGCGCTGTGCAAGTGTGTTGG  
AGACCCGACGTAGACAAACCAAGTGTAA

>KU298930.1:102-551 Human papillomavirus type 67 isolate 87B.67, complete genome

Ao colar uma sequência, pressione "ENTER" no seu teclado, assim o programa reconhece que as informações a seguir serão de uma outra sequência.

Em seguida salve o arquivo no diretório escolhido.

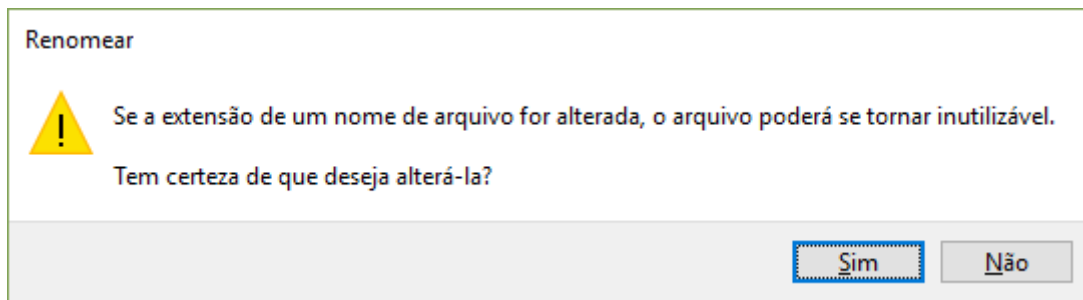
Para usar os dados que você coletou, o arquivo em formato ".txt" (salvo em bloco de notas) deve ser alterado para formato FASTA. Para fazer isto basta apenas renomear o arquivo e modificar a extensão do arquivo para ".fas", como nas imagens abaixo:



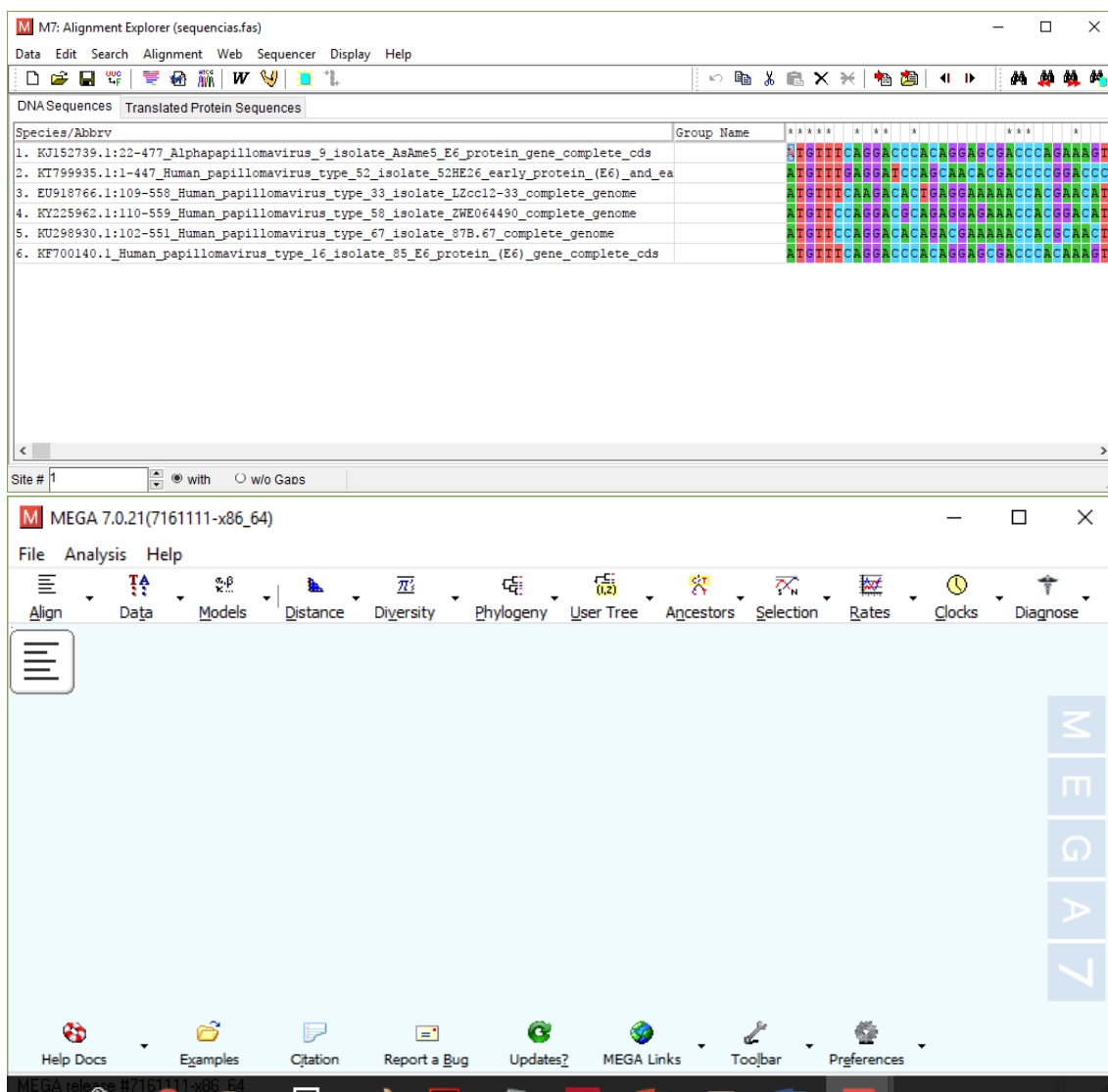




Um aviso como este pode aparecer, clique em “sim” e aguarde a conversão.

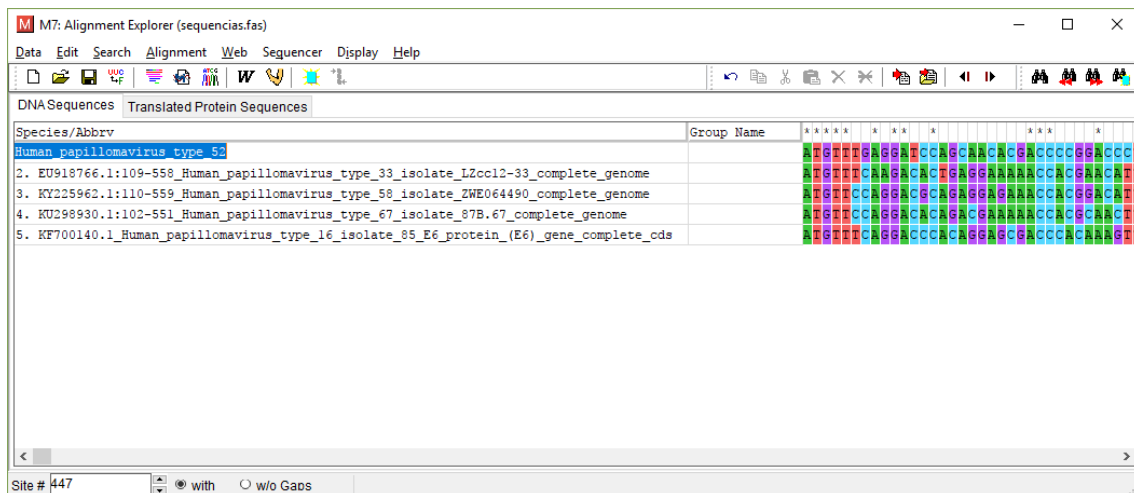


Agora seu arquivo está pronto para ser utilizado pelo programa MEGA 7, basta abri-lo com dois cliques, e duas telas surgirão:

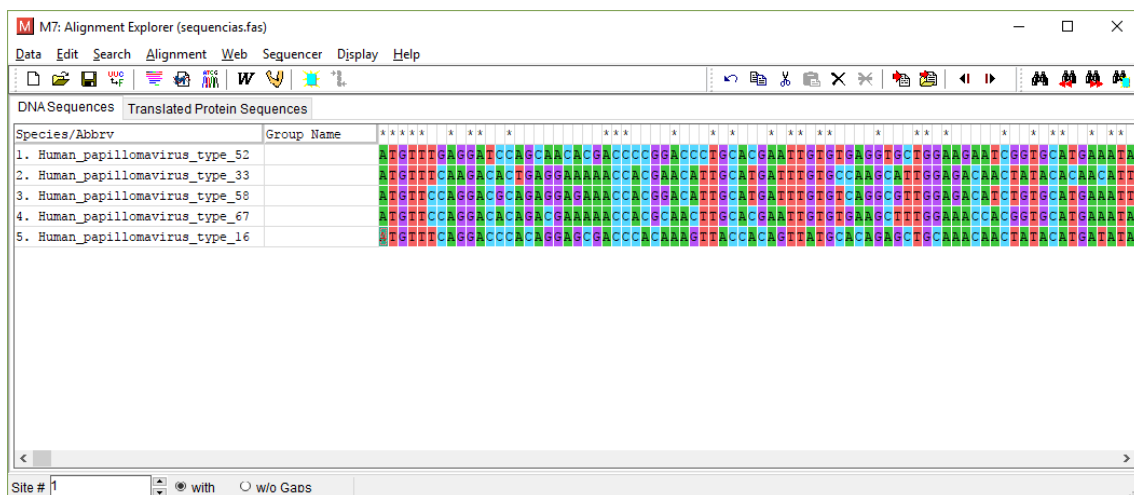


Caso ache necessário, você pode renomear suas sequências, facilitando a análise no final, para realizar esta operação, basta clicar duas vezes sobre o nome da sequência e editá-la.





Mantenha apenas o nome do vírus e o tipo, repita este processo para todas as sequências.



Para realizar o alinhamento o programa oferece dois métodos, o método *ClustalW* e o *MUSCLE*.

### ***ClustalW***

Este método, realiza um sequenciamento par a par e a partir deste, cria uma árvore guia. Esta árvore, é utilizada para alinhar as sequências restantes com o primeiro alinhamento formado pelo programa. O problema deste método se dá principalmente quando as sequências são pouco similares, enviesando a árvore guia e o alinhamento em si.

### ***MUSCLE***

Este método realiza processos muito semelhantes ao descrito acima, porém o MUSCLE realinha as sequências e recalcula as pontuações para cada alinhamento refeito, criando assim o que chamamos de refinamento dos alinhamentos.







Neste caso, pode-se optar pelo método *MUSCLE*, ao clicar nele o programa questionará se você quer alinhar a sequência de DNA ou códons, para esta análise, vamos alinhar os códons. Outra maneira de realizar este procedimento, é traduzindo a sequência de nucleotídeos e alinhando as sequências traduzidas.

**M7: MUSCLE - AppLink**

Option	Selection
<input type="checkbox"/> Presets	None
<b>Gap Penalties</b>	
Gap Open	-2.9
Gap Extend	0
Hydrophobicity Multiplier	1.2
<b>Memory/Iterations</b>	
Max Memory in MB	1530
Max Iterations	8
<b>More Advanced Options</b>	
Clustering Method (Iteration 1,2)	UPGMB
Clustering Method (Other Iterations)	UPGMB
Min Diag Length (lambda)	24
Genetic Code (when using cDNA)	Standard
Alignment Info	MUSCLE Citation: Edgar, Robert C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32(5), 1792- 1797.

Ao clicar no MUSCLE, clique em “compute” e aguarde o resultado do alinhamento.

M7: Alignment Explorer (sequences - Copia.fas)

Data Edit Search Alignment Web Sequencer Display Help

DNA Sequences Translated Protein Sequences

Species/Abb/Group Name

1. Human\_pa

2. Human\_pa

3. Human\_pa

4. Human\_pa

5. Human\_pa

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984





OBS.: Em alguns casos códons de parada podem aparecer no meio da sequência traduzida, neste caso é válido verificar se o gene em questão é traduzido de acordo com o código padrão de a.a. ou se é traduzido por um código variado. Essa análise só será necessária se a sequência analisada for codificante.





## Tutorial Filogenia

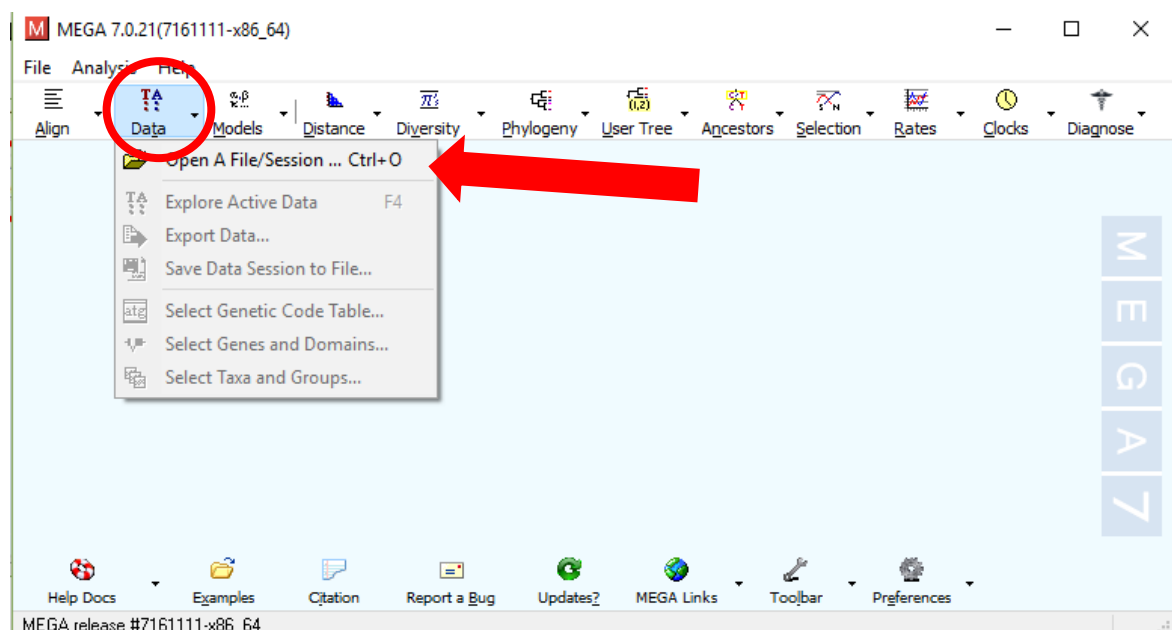
Hoje vamos usar o alinhamento de sequências múltiplas (MSA) o qual geramos a partir da última aula para inferir as relações evolutivas entre as sequências. As análises filogenéticas produzem diagramas com ramificações que podem ilustrar claramente as relações entre sequências que não são aparentes a partir de análises como alinhamento múltiplo de sequências (MAS). As árvores filogenéticas são, obviamente, úteis para estudos evolutivos e comparativos com foco em relações e padrões de divergência evolutiva, mas eles também estão se tornando cada vez mais importantes para a geração de hipóteses sobre se o gene ou a proteína funcionam para estudos moleculares e bioquímicos.

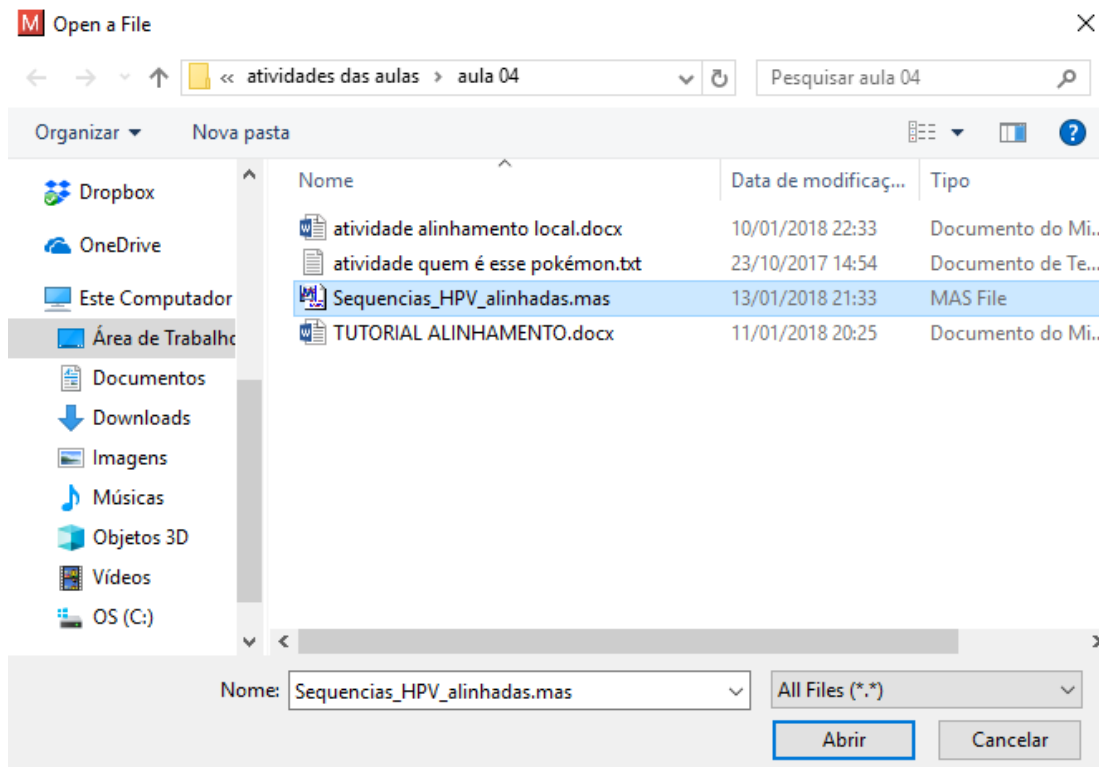
### Neighbour-joining

O Neighbour-joining é rápido e bastante confiável, e, portanto, pode ser utilizada para fazer árvores iniciais possam construir hipóteses sobre a topologia da árvore antes de passar para os programas mais rigorosos. Neste tutorial, iremos focar na construção filogenética utilizando este método e o método de máxima parcimônia, comparando seus resultados.

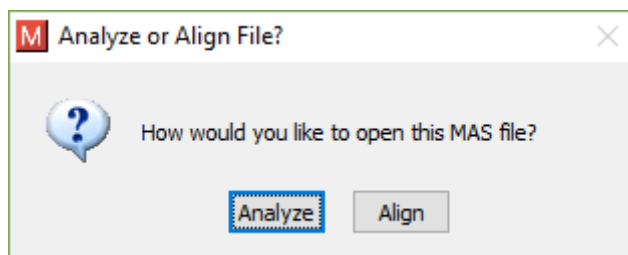
Primeiramente localize as sequências alinhadas que foram salvas na aula anterior: (HPV gene E6).

Em seguida, abra o MEGA, clique em “TA Data”; “Open a file” e selecione o arquivo das sequências alinhadas.

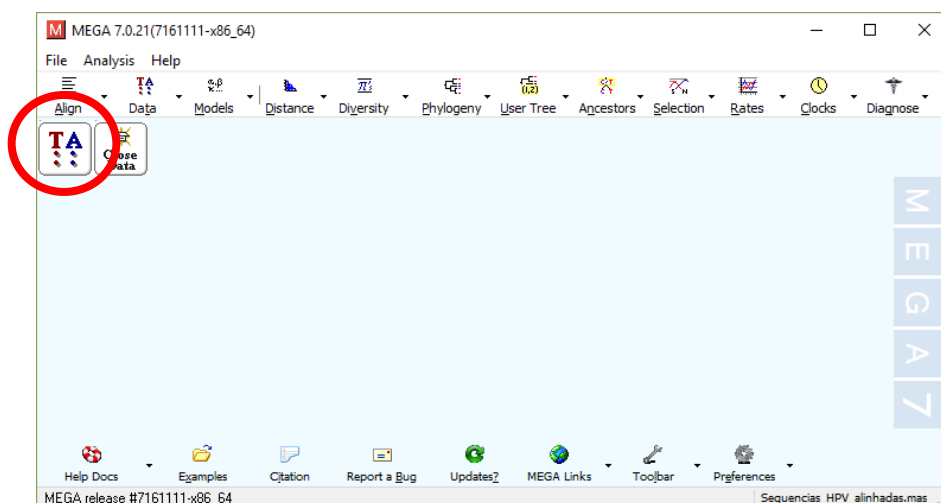




Em seguida clique em “Analyze”.



Agora você pode obter algumas informações acerca da sua sequência clicando no “TA” que apareceu na interface do programa.





Ao clicar neste ícone, uma nova janela se abrirá onde você poderá obter informações como sítios variáveis e conservados, tradução em proteínas, quem são as trincas de cada códon etc...

Sítios conservados

Sítios variáveis

Traduzir em aminoácidos

Name	Sequence
1. Human_papillomavirus_type_33	A T G T T T C A A G A C A C T G A G G A A A A A C C A C G A A C A T T G C A T G A T T T G
2. Human_papillomavirus_type_52	A T G T T T G A G G A T C C A G C A C A C G A C C C C G G A C C C T G C A C G A A T T G
3. Human_papillomavirus_type_58	A T G T T C A G G A C G C A G A G A A A A C C A C G G A C A T T G C A T G A T T T G
4. Human_papillomavirus_type_67	A T G T T C A G G A C A C A G A C G A A A A C C A C G C A A C T T G C A C G A A T T G
5. Human_papillomavirus_type_16	A T G T T T C A G G A C C C A C A G G A G C G A C C C A C A A A G T T A C C A C A G T T A

As trincas são representadas por um retângulo que aparece em cima de cada três colunas:

Primeira trinca de nucleotídeos

Segunda trinca de nucleotídeos

Quarta trinca de nucleotídeos

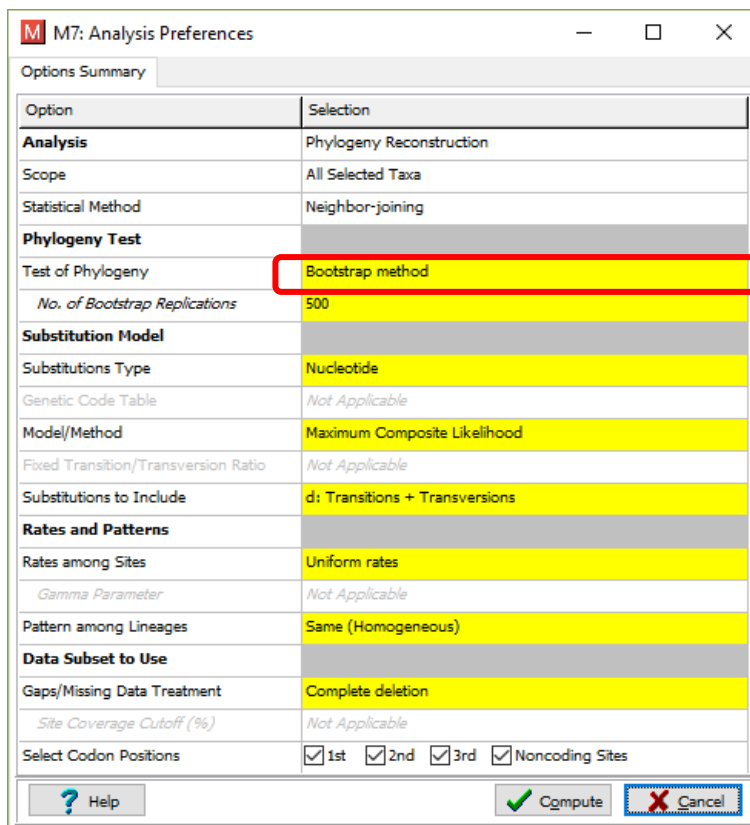
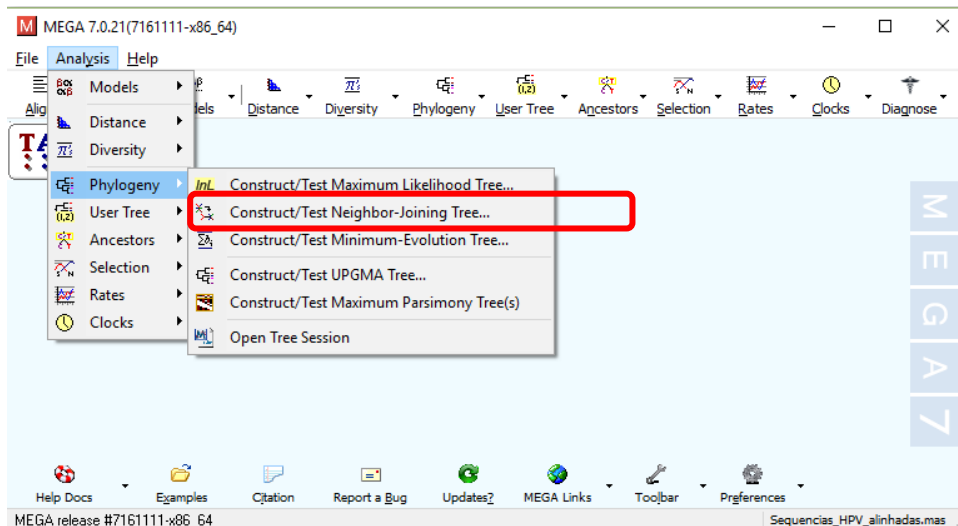
Name	Sequence
1. Human_papillomavirus_type_33	A T G T T T C A A G A C
2. Human_papillomavirus_type_52	A T G T T T G A G G A T
3. Human_papillomavirus_type_58	A T G T T C C A G G A C
4. Human_papillomavirus_type_67	A T G T T C C A G G A C
5. Human_papillomavirus_type_16	A T G T T T C A G G A C





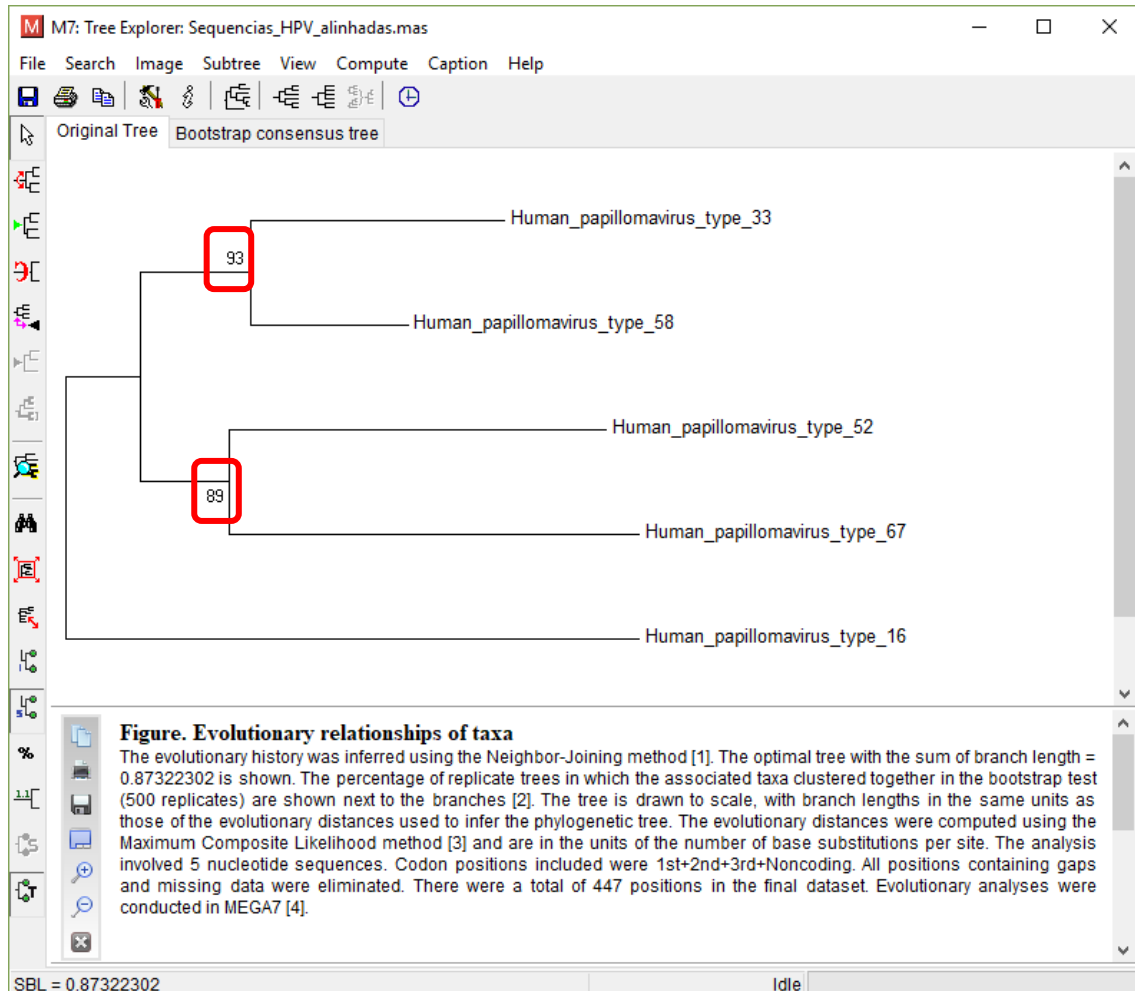
Presumindo-se que as sequências já estão alinhadas podemos iniciar a construção da filogenia deste grupo.

Na janela principal do MEGA clique em: **Analysis > Phylogeny > Construct/Test Neighbor-Joining Tree**. Vamos ficar com os parâmetros padrões, mas certifique-se de escolher um: "**método Bootstrap**", sob Phylogeny Test/ Test of Phylogeny. Clique em compute.





Seu retorno deve ser uma árvore filogenética bem formatada em uma nova janela. Observe os valores acima de cada nó ou ramo na árvore. Estes são chamados de valores de bootstrap, e são uma medida de confiança estatística para cada nó. Qualquer pontuação de bootstrap  $\geq 70$  é geralmente considerada bastante confiável.



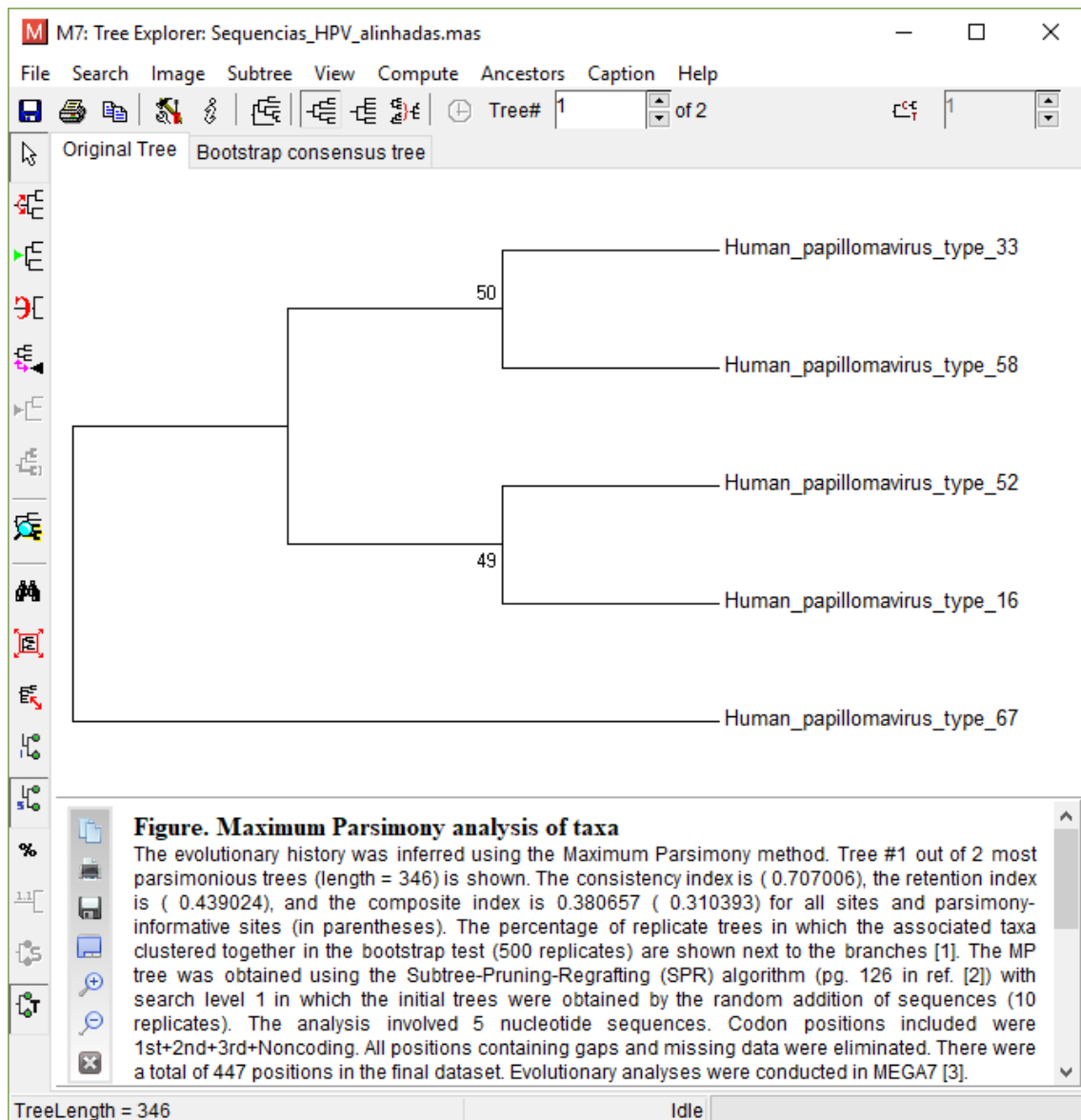
1 - Quantos nós confiáveis de bootstrap existem?

2- Você tem muita confiança nesta árvore?

Vamos tentar agora através do método de Máxima Parcimônia, lembre-se de selecionar o bootstrap como teste para a filogenia:

Após estes passos você possivelmente obterá uma árvore semelhante a esta:

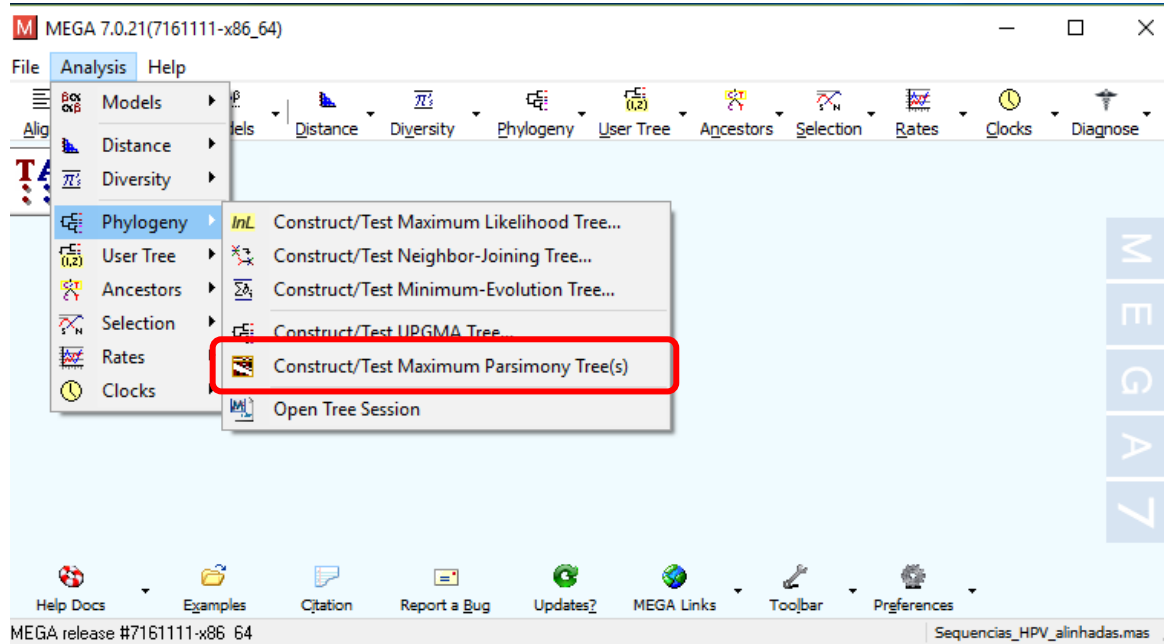




- 1 - Quantos nós confiáveis de bootstrap existem?
- 2- Você tem muita confiança nesta árvore?
- 3 – Houve alguma alteração entre as árvores realizadas?
- 4 – Em qual das duas árvores você teria mais confiança? Porquê?
- 5 – Existe uma maneira de melhorar a sua hipótese de “melhor árvore”?







# **Bioinformatics for undergraduate students of Biological Science course under active methodology**

Everton Silva Mota and Marcus Vinícius de Aragão Batista

[A ser submetido na revista Biochemistry and Molecular Biology Education]

## **Abstract**

With the advancement of Molecular Biology, several datasets have emerged and together with them, there was a need to systematize and analyze this data for the generation of information. To perform such a function arises a new area that encompasses the Biological Sciences, Mathematics and Computer Science: The Bioinformatics. Few Brazilian universities offer courses that deal with aspects related to bioinformatics at the undergraduate level, thus, many biologists leave universities without having any knowledge about Bioinformatics. Because it is an area closely linked to Molecular Biology, Bioinformatics proves to be an important tool to aid in the teaching of some disciplines such as Genetics. In this context, the present work aims to introduce the discipline of Bioinformatics for undergraduates of the Biological Sciences course enrolled in the discipline of Molecular Genetics, through an active teaching methodology, aiming to evaluate the relevance and importance of this discipline for undergraduates, as well as the effectiveness of this teaching methodology. It was verified that the students themselves feel the need for a more in-depth approach to Bioinformatics in the academic curriculum, besides; it is observed that active methodologies assist in the teaching-learning process of these students, facilitating a meaningful learning.

**Key words:** Teaching of Biology; Molecular genetics; Active methodology; Curricular Innovation.

## **Introduction**

Since the discovery of DNA, several researches have been contributed with more information about this molecule; these researches have brought important data explaining how DNA acts in organisms, and how it is possible to manipulate it in laboratories. From this information, new equipment has emerged, allowing more in-depth studies.

With the advance of laboratory equipment's, there was also the advancement of Molecular Biology, thus increasing the data and information about it. With the number of these data growing, then the need arises to organize and analyze them, effectively generating information. For this issue, appears the Bioinformatics, that tries to solve these questions. (Farah, 2007; GenBank, 2017; GOLD, 2017).

In Brazil, undergraduate courses in biological sciences, for example, hardly offer to undergraduate students some knowledge about bioinformatics, nor do they use these tools to facilitate the teaching-learning process in undergraduate classes.

In Brazil, the need to generate human resources capable of acting in this area is noticeable, since the increase in molecular data generated by scientific researches establishes a significant demand. However, many undergraduate students leave universities without even knowing what bioinformatics is about. Ribeiro Junior; Oliveira; Ceccatto (2012), show this fact when they say that the insertion of bioinformatics is very rare and still occurs generally in postgraduate courses, with only a few universities introducing bioinformatics knowledge in undergraduate studies.

It is known that in the academic and social environment the technique is strongly associated with the theory and the opposite is also valid (3). Therefore, the professor might to seek always to associate theories with the practices in order to help the learning process.

As (4) says, the professors should provide some characteristics to the student, these would be:

- Acquisition of creative, participative and questioning awareness;
- Present theoretical references for analysis and interpretation of reality;
- Be able to link theory and practice.

The choice of methodologies that facilitate this process is of great value for the reflection of the professor, so that these reflections help the students to follow the path to the objective proposed by the professor for the course.

The active methodology of teaching is a viable alternative to promote such characteristics, and those that enable the student to build their knowledge, while becoming the center of the discussions related to the theme define it. (5).

Bioinformatics brings the opportunity of a differentiated pedagogical approach through problem solving (6). Using your tools to stimulate the student in the search for solutions is an interesting method to develop skills such as those cited by Silva, (2012).

It is necessary to mention that bioinformatics is a growing area and very important for biologists, however, there is a lag in the teaching of this in undergraduate courses, so this study aims to evaluate how the introduction of the Bioinformatics interferes in the quality of student training of Biological Sciences.

Several authors have already used bioinformatics in order to improve the training of undergraduates. In (7), students used programming language to understand processes related to Molecular Biology such as: Transcription, Translation, Complementarity of the DNA tape, and others. The work brought a differentiated approach to the teaching of Molecular Biology.

For (8), Bioinformatics has shown promise for the teaching of proteomics, so that even with the aid of tutorials, students are more active in the process and in fact think in an analytical way.

According to (9), the active methodology and Bioinformatics can be used to improve students' writing and reasoning skills. In their work, the authors propose to students to write a project based on previously provided biochemistry courses, but in these courses the contents are mediated by Bioinformatics tools under active methodology, facilitating the understanding of the processes approached. Thus, the student can then think of ways to use the tools studied for the creation of other research projects.

Other works also used Bioinformatics to assist in the teaching process of several disciplines, such as: Mathematics (6,10) , Biochemistry (8,9,11,12) Programming, Chemistry, Physics (13), Physiology (7,14), among others.

Therefore, under an analysis of the current situation of bioinformatics, especially in Brazil, the need to introduce of Bioinformatics in the academic curriculum of the undergraduate courses in Biological Sciences is questioned. It is still valid to inquire that the use of bioinformatics tools in an active way contributes significantly with the learning of undergraduate students and consequently in the training of these future professionals.

Thus, this study aims to introduce the Bioinformatics to undergraduates of the Biological Sciences Course through an active pedagogical methodology, where the students will seek solutions of biological problems proposed by the professors, thus, we aim to verify if such problems presented amplify students' investigative skills. In addition, it is sought to observe from the perspectives of the students the importance of this discipline in the curriculum of the course.

## **Material and Methods**

The present study was carried out at the Federal University of Sergipe, where the sample were undergraduate students who accepted to participate in the proposed work. The proposal was made for a group of Molecular Genetics composed of 27 students where 100% of the students studied Biological Sciences.

This study was submitted to the Research Ethics Committee Involving Human Beings, through the Brazil Platform, obtaining approval on 11/12/2017 (CAAE: 80239717.3.0000.5546). The participants, individually, received the questionnaire with the attached free informed consent form, according to resolution 510/2016.

Were planned lessons in order to introduce the main aspects of Bioinformatics as well as to present the tools used, as shown in table 01. Were elaborated theoretical and practical lectures for all the classes that involved the use of computational tools, where students understood the theory of each tool and put this knowledge into practice. For the accomplishment of the practical procedures the students used tutorials and scripts elaborated specifically for this purpose. Thus the student had the opportunity to know more about each tool that would be used.

All classes were planned with the purpose of creating the best possible scenario so that the students could then explore the tools in an autonomous way, discussing in groups and acting in an active way on the problems presented to them.

In all, four tutorials and one script were elaborated, the tutorials being used for the sequence analysis classes, BLAST alignment, multiple alignments, and phylogeny, and the script was prepared for application in the database class.

Table 01: Table with the themes of the classes, hours dedicated to each theme.

ACTIVITY	HOURS
Introduction to Bioinformatics	2
Chromatograms analysis	2
Database	2
Sequence alignment	4
Molecular Phylogeny	4

Questions	4
Presentation of problem solutions	2

In addition to the theoretical-practical classes, four classes were available for students to solve doubts related to the operation of the software that implied in solving their problem.

The theoretical classes sought to address the main characteristics of the subject in question and demonstrate the functioning of the programs that would be used. After classes, students solved a problem of low complexity in the classroom with the aid of a tutorial and the professor, in this way the students became familiar with the programs used. After all the theoretical and practical classes, the students formed five groups and participated in a lottery that defined their themes and biological problems to be solved.

Each problem would have some questions that should be answered and the students should prepare a presentation where they would demonstrate their problems, their results and the methodologies used

Each group was responsible for finding solutions to the problems presented to them in the way they believed to be the best possible.

It is observed that the purpose in establishing problems is to stimulate the student in seeking the possible solutions for such in an autonomous way, aiming at amplifying the critical thinking, the reasoning and investigative abilities of the students.

Each group had the theme of one of the classes, and their respective problems, as shown in the table below:

Table 02: Distribution of the themes to the groups formed.

Group	Theme
Group 01	Chromatograms analysis
Group 02	Database
Group 03	Sequence alignment with BLAST
Group 04	Multiple alignment
Group 05	Phylogeny

The Issues of each group were as follows:

### **Issue 01: Sequencing analysis**

You are part of a bioinformatics lab team, and have received from another lab some results from a Sanger-like sequencing that was done. The said laboratory requests its services in order to answer the following questions:

- a) Did all the samples show a good sequencing? Justify.
- b) What is the cut-off value of PHRED used in this analysis? why?

Are the contigs formed correctly? What is the size obtained in the consensus sequence?

### **Issue 02: Databases**

Your professor is requesting a set of information regarding the Human Papillomavirus E6 (HPV) gene and requests the following responses:

- a) Which database will you use to access the information of a single gene? It's because?
- b) What is the size of the gene, the location of the gene in the genome of the organism and how many amino acids does the protein have?
- c) What is the function / functions of the gene in question?

### **Issue 03: Alignment with BLAST**

A research group divided the tasks for the execution of a work, you and your team were responsible for identifying which organisms belong to a set of sequences, finding species genetically similar to the samples that were sent to you through a local alignment search tool. They asked for the following answers:

- a) There are 5 ways to perform searches according to the sequences you get, knowing this, what method will you use (BLASTn, BLASTp, tBLASTn ...)? Justify.
- b) Are the results found in your search statistically significant? why?

## Issue 04: Multiple Alignment

Upon receiving a set of sequences, your advisor asks you to perform an alignment and do a simple analysis on these, and asks you to check:

- a) Which method would you choose to perform the alignment? ClustalW or MUSCLE? Why?
- b) After the alignment it was possible to verify that at the 4th site there is a substitution of bases, does this substitution change the corresponding amino acid? What do we call this type of mutation?
- c) At site 634 is it possible to observe that a Gap was inserted into the *Tarsius syrichta* species, that Gap modified the sequence reading matrix? If yes, how to minimize the effects of the insertion of this Gap taking into consideration the reading matrix?

## Issue 05: Phylogeny

Among the tasks divided into your research group, you and your team were given the function of performing the phylogeny of some aligned genes. To carry out the phylogenetic construction your boss asks you the following:

- a) Among the methods of phylogenetic tree construction, which one will you use and for what reason? (Neighbor-joining, maximum likelihood, maximum parsimony)
- b) Is this tree reliable? Can it explain the phylogenetic relationships between the species analyzed? why?

The collection was done through a questionnaire of 11 questions, where 8 are subjective questions, aiming to obtain greater clarity in the answers, and 3 are objective facilitating the categorization of the data. The data obtained were tabulated (for objective questions) and analyzed individually (for subjective questions).

## Results and discussion

A total of 23 questionnaires were received. Among the 47% (11 participants) answered the first question in a similar way, highlighting the importance of introducing



Bioinformatics in undergraduate courses, as well as the lack of information about this topic, this can be observed in some of the answers below:

Student 01: "Important. Because we have little contact with this kind of subject, even more as practices. "

Student 02: It is interesting because it is something that we do not have in the curriculum and it is a very used area in the area of research and laboratory. "

Student 03: "I found it important because it introduced knowledge beyond the curriculum, opening up new perspectives for other forms of employment in the field of biology."

Student 04: "I found it very interesting, because it is an area that is not widely spoken in undergraduate courses and is currently widely used in postgraduate studies [...]"

Student 05: "To great advantage, because we do not have this discipline in the curriculum of our course."

Based on these quotations, it is notable that students realize that there is a need to use Bioinformatics even in the undergraduate curriculum, even though in a more superficial way. Thus, the professor can contribute with aspects that aid in the learning of these students, basic knowledge about this subject is likely to be fundamental for the postgraduate research of many of these students.

According to Wightman e Hark, (2012), the introduction of bioinformatics concepts to students early in their careers is effective and results in significant learning gains. This was clearly demonstrated in the students' own speeches.

The other students summarized their ideas regarding this question, arguing that the introduction was very useful for a better understanding of laboratory analyzes.

The second question asked participants if bioinformatics classes would expand their investigative skills. 82% (19 participants) said they believed that their investigative abilities were indeed broadened according to the content approach.

The students' evaluation of the course in general was very positive, where 73% (17 participants) evaluated how great and the rest as good.

These two issues can demonstrate that the introduction of this discipline has contributed important aspects to the professional future, establishing critical and analytical thinking has not been an easy task in teaching and the use of these tools under this approach can contribute to this purpose.

Regarding the importance and relevance of Bioinformatics in the undergraduate curriculum in Biology, 22 participants consider it important for their training. In addition, all the students agreed that in fact this subject is relevant to the curriculum, however, some attributed greater importance of this discipline to while others (56%) said that they were able to increase the number of trainees and to prepare them more for the job market.

One of the fundamental points for this work is the methodology used to approach the chosen themes, since it used praxis in the search of assigning meaning to the fundamental concepts of Bioinformatics. The participants of the survey were also invited to evaluate this question through questions 6 and 7.

With regard to this issue, 100% of the participants considered this approach to be more efficient than the expository only approach, some even justify the reason why this methodology was well evaluated by them as can be observed below:

Student 01: "It is more efficient, since it makes the student" protagonist ", he who will seek and build the knowledge within the subjects that are exposed"

Student 04: "Very efficient, because it stimulates the student."

Student 06: "More efficient, therefore, made us exercise what we learned in theory and thus know how to identify in what kind of situation we can apply it in practice."

Student 07: "Much more effective, because if we only worked the exhibition without the active search, we would see only closed problems that would not represent real problems that we could face in our careers.

Research stimulates creativity and curiosity and increases the critical attitude of students, who develop a new attitude towards knowledge and become more motivated to learn (Rodrigues e Silva, 2009; Zompero e Laburú, 2010; Munford e Lima, 2013).

In relation to the suggestions for approaching the subject, some students brought interesting characteristics that should be taken into account, for example: "to follow a step

by step from the collection of the material to the analysis", or "every student to think about problems and find ways to solve it."

These statements connote how students become enthusiastic when feeling part of the teaching-learning process, as well as feel more protagonists when actively participating in the construction of their knowledge.

As for the exclusive approach of the subject, the class was divided; some believe that with an exclusive approach there would be the possibility of further deepening of the contents, besides more time for the accomplishment of the activities. Others believe that interdisciplinarity may be the key to an efficient approach, since Bioinformatics is used in many areas.

Regarding previous knowledge about Bioinformatics, 12 participants claimed not to know the discipline before this approach the others, said they knew the subject superficially.

## **Conclusion**

With subsidy in the above, it is noticeable that few students know Bioinformatics, and those who know often do not know how to use their tools or in what areas they apply. Thus, the importance of introducing the tools of Bioinformatics in the undergraduate curriculum in Biological Sciences is notorious, since it has proved to be a facilitator of the teaching-learning process of the students, besides expanding skills considered fundamental for these professionals and above all stimulating the critical thoughts about the subjects explored.

The participants of the research report themselves to be able to identify that in fact this discipline helped them in their perception of possible solutions to a problem. In addition, it brings Bioinformatics to undergraduate studies is of extreme importance for the academic training of undergraduates, generating human resources trained not only for the immediate job market but also for the endowment in graduate studies.

Thus, it is possible to infer that the application of activities such as those proposed in this work adds knowledge and amplifies important skills for the professional future, which demonstrates how these methodologies improve the quality of biologist training.

## References

1. Farah SB. DNA segredos e mistérios. Sarvier. São Paulo; 2007. 1-538 p.
2. Ribeiro Junior HL, Oliveira RTG de, Ceccatto VM. Bioinformática como recurso pedagógico para o curso de ciências biológicas na Universidade Estadual do Ceará – UECE – Fortaleza, Estado do Ceará. Acta Sci Educ [Internet]. 2012;34(1):129–40. Available from: <http://periodicos.uem.br/ojs/index.php/ActaSciEduc/article/view/14584>
3. Ribeiro D da S, Jesus TF de. A Práxis Pedagógica no Ensino Superior: um relato baseado na necessidade de ações inovadoras de avaliação. Cairu em Rev. 2015;4(Jan/Fev nº05):124–42.
4. Silva ACR. Educação por competência. 2012;
5. Melo B, Sant’Ana GA. A Prática da Metodologia Ativa: Compreensão dos discentes enquanto autores do processo ensino-aprendizagem. EscsEduBr. 2013;23(4):327–39.
6. Wightman B, Hark AT. Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. Biochem Mol Biol Educ. 2012;40(5):310–9.
7. Nunes R, Barbosa de Almeida Júnior E, Pessoa Pinto de Menezes I, Malafaia G. Learning nucleic acids solving by bioinformatics problems. Biochem Mol Biol Educ. 2015;43(5):377–83.
8. Badotti F, Barbosa AS, Reis ALM, do Valle ??talo Faria, Ambr??sio L, Bitar M. Comparative modeling of proteins: A method for engaging students’ interest in bioinformatics tools. Biochem Mol Biol Educ. 2014;42(1):68–78.
9. Mertz P, Streu C. Writing throughout the biochemistry curriculum: Synergistic inquiry-based writing projects for biochemistry students. Biochem Mol Biol Educ. 2015;43(6):408–16.
10. Chapman BS, Christmann JL, Thatcher EF. Bioinformatics for Undergraduates. Biochem Mol Biol Educ. 2010;34(3):180–6.
11. Inlow JK, Miller P, Pittman B. Introductory bioinformatics exercises utilizing hemoglobin and chymotrypsin to reinforce the protein sequence-structure-function relationships. Biochem Mol Biol Educ. 2007;35(2):119–24.
12. Junior HLR. Atividades práticas de bioinformática como recurso pedagógico para a formação de professores do curso de ciências biológicas na Universidade Estadual do Ceará - UECE Fortaleza - ceará. 2010.
13. Furge LL, Stevens-Truss R, Moore DB, Langeland JA. Vertical and horizontal integration of bioinformatics education: A modular, interdisciplinary approach. Biochem Mol Biol Educ. 2009;37(1):26–36.
14. Zhang X. Using arabidopsis genetic sequences to teach bioinformatics. Biochem Mol Biol Educ. 2009;37(1):16–23.